REMARKS

Claims 1-5 and 7-11 are presented for examination. Applicants respond below to the specific rejections raised by the Examiner in the pending Office Action. For the reasons set forth below, Applicants respectfully traverse.

35 U.S.C. §§ 102(e) and 103(a) – Afar *et al.*

The Examiner has rejected claims 1-5 and 7-11 under 35 U.S.C. § 102(e) or alternatively 35 U.S.C. § 103(a) over Afar *et al.* (U.S. Pub. No. 2005/0019870). The Examiner states that Afar discloses a prostate tumor associated gene (24P4C12) and its encoded protein. The Examiner asserts that the 24P4C12 protein, which is 710 amino acids long, is identical to amino acids 9-321 of SEQ ID NO: 10. The Examiner states that Afar discloses and claims anti-24P4C12 antibodies.

The Afar reference does not expressly or inherently disclose the claimed antibodies

In response to Applicants' arguments that the Afar reference does not expressly or inherently disclose an antibody which specifically binds to the polypeptide of SEQ ID NO:10 since the majority of 24P4C12 does not correspond to SEQ ID NO:10, the Examiner asserts that "[b]ased on the Kyte-Doolittle plot, Afar clearly identifies immunogenic regions that are identical to the immunogenic regions in applicants' SEQ ID NO:10 and clearly discloses antibodies to those identical immunogenic regions." *Office Action* at 5 (emphasis added). To support this assertion, the Examiner cites page 34, full paragraph 3. *See Office Action* at 6. No antibody is disclosed at "page 34, full paragraph 3." Page 34 of US 2005/0019870 only contains a portion of the sequence listing for the application, and there is no disclosure of any specific antibody on that page.

Applicants have reviewed the Afar and find no disclosure in Afar that "clearly identifies immunogenic regions that are identical to the immunogenic regions in applicants' SEQ ID NO:10 and clearly discloses antibodies to those identical immunogenic regions." The closest disclosure Applicants can find is in ¶[0112], where Afar states that "[t]he amino acid sequence of the 24P4C12 ... may be used to select specific regions of the 24P4C12 protein for generating antibodies," and that these regions can be identified using methods known in the art, such as Kyte-Doolittle analysis. *Afar* at ¶[0112] (emphasis added). However, Afar does not disclose

that Afar et al. actually utilized the 24P4C12 sequence to select a specific region for making antibodies (with one exception discussed below). As a result, Afar does not identify or disclose any specific hydrophilic or immunogenic regions of the 24P4C12 protein that can or should be used. Thus, contrary to the Examiner's assertion, Afar does not "clearly identif[y] immunogenic regions that are identical to the immunogenic regions in applicants' SEQ ID NO:10 and clearly disclose[] antibodies to those identical immunogenic regions." At best, Afar suggests making antibodies to immunogenic regions of the 24P4C12 protein – not a particularly illuminating disclosure since one cannot make antibodies to non-immunogenic regions. Afar's teachings are general in nature and cannot be anticipatory because these teachings do not describe the invention in as full detail as that which is claimed. Absent such teachings, Afar cannot anticipate the claims.

At best, this is a disclosure of a potential genus of antibodies that one of skill in the art could make, and no more. It is not an express or inherent disclosure of any particular species of antibodies, much less an antibody as claimed in the present application. Afar simply does not disclose a particular species of antibody based on this teaching. Applicants invite the Examiner to point to the particular portion of Afar where a particular antibody is disclosed – Applicants are aware of only one.

The only example of an antibody to a specific portion of the 24P4C12 protein disclosed in Afar is found in Example 4. Afar teaches a polyclonal antibody to a peptide "corresponding to amino acids 1-14 ... of the 24P4C12 protein sequence." *Afar* at ¶[0235]. As Applicants have previously noted, according to the Examiner, amino acids 398-710 of 24P4C12 disclosed in Afar are identical to amino acids 9-321 of SEQ ID NO:10, but not amino acids 1-397. Thus, the Examiner has not established that amino acids 1-14 of the 24P4C12 protein sequence are similar to SEQ ID NO:10 of the instant specification. Absent such a showing, there is no basis to assert that the antibodies disclosed in Example 4 of Afar anticipate or render obvious the pending claims.

In addition, Afar discloses a prophetic example of monoclonal antibodies to the 24P4C12 protein in Example 9. That example discloses that "[i]n order to generate 24P4C12 monoclonal antibodies, a glutathione –S-transferase (GST) fusion protein encompassing a 24P4C12 protein is synthesized and used as immunogen. Alternatively, 24P4C12 can be conveniently expressed in 293T cells transfected with a CMV-driven expression vector encoding 24P4C12 with a C-

terminal 6xHis and MYC tag (pcDNA3.1/mycHIS, Invitrogen)." *Afar* at ¶[0251] (emphasis added). As noted previously, according to the Examiner's alignment, 397 of the 710 amino acids of 24P4C12 are apparently completely different from SEQ ID NO:10. That means that 56% of 24P4C12 differs from SEQ ID NO:10. Thus, there is no reason to believe that a monoclonal antibody generated by use of the entire 24P4C12 protein would satisfy the limitations of the pending claims.

In sum, Afar does not disclose any antibodies to any specific region of 24P4C12, other than amino acids 1-14. Instead, there is merely a generic disclosure of antibodies to 24P4C12. While it is <u>possible</u> that an antibody to 24P4C12 would bind the polypeptide of SEQ ID NO:10, it is not a certainty since more than half of 24P4C12 apparently bears no similarity to SEQ ID NO:10. Mere possibility is not sufficient for inherent anticipation: "Inherency, however, **may not be established by probabilities or possibilities**. The mere fact that a certain thing **<u>may</u>** result from a given set of circumstances **is not sufficient**." *M.P.E.P.* §2112 ¶IV (8th ed. 2004), quoting *In re Robertson*, 169 F.3d 743, 745 (Fed. Cir. 1999) (emphasis added).

The Examiner attempts to overcome this shortcoming by generating a Kyte-Doolittle plot some 5 years after the filing of the instant application. This plot is in no way inherent in Afar, and does not constitute part of the Afar disclosure. Thus, regardless of whether the plot is accurate or not, a point Applicants have not conceded, it is not a part of the Afar disclosure and does not transform Afar's generic disclosure of possible antibodies to 24P4C12 into a disclosure of specific antibodies. As noted previously, the plot shows that there are hydrophilic regions throughout the entire length of 24P4C12, <u>including amino acids 1-397</u>, which have no apparent similarity to SEQ ID NO:10. Thus, the Examiner's post-filing Kyte-Doolittle plot, even if accurate, discloses nothing more specific regarding the claimed antibodies beyond Afar's generic disclosure of generating antibodies. Absent the plot, Afar generically discloses possible antibodies to immunogenic portions of 24P4C12, some of which antibodies <u>may</u>, but do not <u>necessarily</u> bind the polypeptide of SEQ ID NO:10. In view of the plot, Afar teaches exactly the same thing – Afar generically discloses possible antibodies to immunogenic portions of 24P4C12, some of which antibodies <u>may</u>, but do not <u>necessarily</u> bind the polypeptide of SEQ ID NO:10. A generic disclosure of possible antibodies to an entire protein is not a disclosure of antibodies to a specific region and the Examiner's plot does not change this fact. The disclosed

generic teachings toward generating antibodies to 24P4C12 clearly do not inherently possess the claimed features of the pending claims.

The Afar reference is not available as prior art

The Examiner stated in the previous Office Action that "[a]ccording to M.P.E.P. § 715.II.B, an affidavit or declaration under 37 C.F.R. 1.131 is not appropriate in situations where the reference U.S. patent or U.S. patent application publication claims the same patentable invention." *Office Action* at 9. Applicants respectfully disagree.

M.P.E.P. § 2136.05 states:

> When a prior U.S. patent, U.S. patent application publication, or international application publication is not a statutory bar, a 35 U.S.C. 102(e) rejection can be overcome by antedating the filing date...of the reference by submitting an affidavit or declaration under 37 CFR 1.131... *In re Mathews*, 408 F.2d 1393, 161 USPQ 276 (CCPA 1969). *M.P.E.P. § 2136.05* (emphasis added).

According to the M.P.E.P., "[a]n exception to the rule arises when the reference is a patent or application published under 35 U.S.C. 122(b) and the reference has claims directed to the same patentable invention as the application claims being rejected. 37 CFR 1.131(a)(1). The reason for this exception is that priority is determined in an interference when the claims interfere." *M.P.E.P. § 2305 I.* However, the M.P.E.P continues, stating that:

> In determining whether a 37 CFR 1.131 affidavit is permitted or not, the examiner should keep the purpose of the exception in mind. If an interference would not be possible at the time the affidavit would be submitted, then the affidavit should be permitted. This situation could arise two ways. ... Similarly, if a published application contains claims to the same invention, but the claims in the published application are not in condition for allowance, then no interference is yet possible. 37 CFR 41.102. Since the claims in the published application might never be allowed in their present form, it is not appropriate to proceed as though an interference would be inevitable. Consequently, an affidavit under 37 CFR 1.131 may be submitted. *M.P.E.P. § 2305 I* (emphasis added).

A non-final Office Action rejecting the claims of the Afar application was mailed on December 28, 2007, rejecting the claims under 35 U.S.C. § 102(e) over several references. A response to the non-final Office Action was forwarded to the Examiner for consideration on May 19, 2008. In view of the fact that the Afar application is not yet in condition for allowance, Applicants submit that it is permissible for Applicants to antedate the Afar reference in the instant case.

The nucleic acid and polypeptide sequences of PRO874 (SEQ ID NOs: 9 and 10 of the instant application) were disclosed by Applicants in US Provisional Application 60/088030 filed June 4, 1998 in Figures 1 and 2, a copy of which is attached as Exhibit 1. Applicants also disclose antibodies to PRO874 in US Provisional Application 60/088030, including polyclonal, monoclonal, and humanized antibodies. *Exhibit 1* at pages 22-28.

The well-established "Stempel Doctrine" stands for the proposition that a patent applicant can effectively swear back of and remove a cited prior art reference by showing that he or she possessed that portion of the claimed invention that is disclosed in the prior art reference. (*In re Stempel*, 113 U.S.P.Q. 77 (CCPA 1957)). In other words, a patent applicant need not demonstrate that he or she made the entire claimed invention in order to remove a cited prior art reference. He or she need only demonstrate prior possession of that portion of his or her claimed invention that is disclosed in the prior art reference and nothing more.

The Examiner asserts that Afar discloses a sequence which is at least partially identical to SEQ ID NO:10, as well as antibodies to the sequence. Applicants demonstrate, by means of the disclosure of antibodies to PRO874 (SEQ ID NO:10) in their provisional application filed June 4, 1998, that they were in possession of as much of the claimed invention as is disclosed in the Afar reference prior to its earliest possible effective filing date, April 12, 1999. Thus, Applicants respectfully submit that Afar is not available as prior art under 102(e).

In light of the above, Applicants respectfully request that the Examiner reconsider and withdraw the 35 U.S.C. §§ 102(e)/103(a) rejection of the pending claims over Afar *et al.*

CONCLUSION

In view of the above, Applicants respectfully maintain that claims are patentable and request that they be passed to issue. Applicants invite the Examiner to call the undersigned if any remaining issues may be resolved by telephone.

Please charge any additional fees, including any fees for additional extension of time, or credit overpayment to Deposit Account No. 11-1410.

Respectfully submitted,

KNOBBE, MARTENS, OLSON & BEAR, LLP

Dated: 5/22/08

By:

Brenden S. Gingrich
Registration No. 60,295
Attorney of Record
Customer No. 30,313
(619) 235-8550

5409856
052108

EXHIBIT 1

A/prov

PROVISIONAL APPLICATION COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION under 37 CFR 1.53(c)

JC533 U.S. PTO 06/04/98

"EXPRESS MAIL" MAILING LABEL
Number: EM104544990 US Date of Deposit: June 4, 1998
I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. 1.10 on the date indicated above and is addressed to the Assistant Commissioner for Patents, Washington, D.C. 20231.
Typed or Printed Name: Geody Domingo

Signed: [signature]

Docket Number:	P-66324/WHD/JPB (G-tech Docket: PR1440)	Type a plus sign (+) inside this box →	+

INVENTOR(s)/APPLICANT(s)

LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
WOOD	WILLIAM	I.	San Mateo, California

TITLE OF THE INVENTION (*280 characters max*)

NOVEL MULTI-SPAN TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME

CORRESPONDENCE ADDRESS

Walter H. Dreger, Registration No. 24, 190
FLEHR HOHBACH TEST ALBRITTON & HERBERT LLP, 4 Embarcadero Center, Suite 3400, San Francisco

STATE	CA	ZIP CODE	94111	COUNTRY	US

ENCLOSED APPLICATION PARTS (*check all that apply*)

X	Specification	*Number of Pages*	42		Small Entity Statement
X	Drawings	*Number of Sheets*	13		Other (*specify*):

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT (*check one*)

X	A check or money order is enclosed to cover the Provisional filing fees		PROVISIONAL FILING FEE AMOUNT ($)	$150
X	The Commissioner is hereby authorized to charge any underpayment or overpayment to Deposit Account Number:	06-1300 (Order No. P-66324/WHD/JPB)		

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

[X] No.

[ ] Yes, the name of the U.S. Government Agency and the Government contract number are: ____________

Respectfully submitted,

SIGNATURE: Jan P. Brunelle — Date: 4 June 1998

TYPED or PRINTED NAME: Jan P. Brunelle — REGISTRATION NO. (*if appropriate*): 35,081

[ ] Additional inventors are being named on separately numbered sheets attached hereto.

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

PATENT
Attorney Docket No.: P-66324/WHD/JPB
(Genentech Docket No.: PR1440)

NOVEL MULTI-SPAN TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME

FIELD OF THE INVENTION

The present invention relates generally to the identification and isolation of novel DNA and to the recombinant production of novel multi-span transmembrane polypeptides designated herein as "PRO874" polypeptides.

BACKGROUND OF THE INVENTION

Membrane-bound proteins and receptors can play an important role in the formation, differentiation and maintenance of multicellular organisms. The fate of many individual cells, e.g., proliferation, migration, differentiation, or interaction with other cells, is typically governed by information received from other cells and/or the immediate environment. This information is often transmitted by secreted polypeptides (for instance, mitogenic factors, survival factors, cytotoxic factors, differentiation factors, neuropeptides, and hormones) which are, in turn, received and interpreted by diverse cell receptors or membrane-bound proteins. Such membrane-bound proteins and cell receptors include, but are not limited to, cytokine receptors, receptor kinases, receptor phosphatases, receptors involved in cell-cell interactions, and cellular adhesin molecules like selectins and integrins. For instance, transduction of signals that regulate cell growth and differentiation is regulated in part by phosphorylation of various cellular proteins. Protein tyrosine kinases, enzymes that catalyze that process, can also act as growth factor receptors. Examples include fibroblast growth factor receptor and nerve growth factor receptor.

Membrane-bound proteins and receptor molecules have various industrial applications, including as pharmaceutical and diagnostic agents. Receptor immunoadhesins, for instance, can be employed as therapeutic agents to block receptor-ligand interaction. The membrane-bound proteins can also be employed for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction.

Efforts are being undertaken by both industry and academia to identify new, native receptor and transmembrane proteins. Many efforts are focused on the screening of mammalian recombinant DNA libraries to identify the coding sequences for novel receptor and transmembrane proteins.

Described herein is the identification and characterization of a novel multi-span transmembrane polypeptide.

SUMMARY OF THE INVENTION

Applicants have identified a cDNA clone that encodes a novel multi-span transmembrane polypeptide, which is designated in the present application as "PRO874".

In one embodiment, the invention provides an isolated nucleic acid molecule comprising DNA encoding a PRO874 polypeptide. In one aspect, the isolated nucleic acid comprises DNA encoding the PRO874 polypeptide having amino acid residues 1 to 321 of Figure 2 (SEQ ID NO:3), or is complementary to such encoding nucleic acid sequence, and remains stably bound to it under at least moderate, and optionally, under high stringency conditions. In other aspects, the isolated nucleic acid comprises DNA encoding the PRO874 polypeptide having amino acid from about X to 321 of Figure 2 (SEQ ID NO:3), where X is any amino acid from about 270 to about 279 of Figure 2 (SEQ ID NO:3), or is complementary to such encoding nucleic acid sequence, and remains stably bound to it under at least moderate, and optionally, under high stringency conditions. The isolated nucleic acid sequence may comprise the cDNA insert of the DNA40621-1440 vector deposited on June 2, 1998, as ATCC ______________ which includes the nucleotide sequence encoding PRO874.

In another embodiment, the invention provides a vector comprising DNA encoding a PRO874 polypeptide. A host cell comprising such a vector is also provided. By way of example, the host cells may be CHO cells, *E. coli*, or yeast. A process for producing PRO874 polypeptides is further provided and comprises culturing host cells under conditions suitable for expression of PRO874 and recovering PRO874 from the cell culture.

In another embodiment, the invention provides isolated PRO874 polypeptide. In particular, the invention provides isolated native sequence PRO874 polypeptide, which in one embodiment, includes an amino acid sequence comprising residues 1 to 321 of Figure 2 (SEQ ID NO:3). Additional embodiments of the present invention are directed to PRO874 polypeptides comprising amino acids X to 321 of Figure 2 (SEQ ID NO:3), where X is any

amino acid from about 270 to about 279 of Figure 2 (SEQ ID NO:3). Optionally, the PRO874 polypeptide is obtained or is obtainable by expressing the polypeptide encoded by the cDNA insert of the DNA40621-1440 vector deposited on June 2, 1998, as ATCC ____________ _____.

In another embodiment, the invention provides chimeric molecules comprising a PRO874 polypeptide or soluble extracellular domain thereof fused to a heterologous polypeptide or amino acid sequence. An example of such a chimeric molecule comprises a PRO874 polypeptide fused to an epitope tag sequence or a Fc region of an immunoglobulin.

In another embodiment, the invention provides an antibody which specifically binds to a PRO874 polypeptide or extracellular domain thereof. Optionally, the antibody is a monoclonal antibody.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A and 1B show a nucleotide sequence (SEQ ID NO:1) containing the nucleotide sequence (SEQ ID NO:2) of a native sequence PRO874 cDNA (nucleotides 1-963), wherein the nucleotide sequence (SEQ ID NO:1) is a clone designated herein as "UNQ441" and/or "DNA40621-1440". Also presented are the locations (as underlined) of oligonucleotide probes based upon the DNA40621-1440 sequence, wherein those probes are designated herein as "40621.tm.f1", "40621.tm.p1" and "40621.tm.r1". The deduced amino acid sequence encoded by the DNA40621-1440 molecule is shown below the nucleotide sequence.

Figure 2 shows the amino acid sequence (SEQ ID NO:3) derived from the nucleotide sequence shown in Figure 1. Also shown are the approximate locations of various other protein domains.

Figures 3A to 3H show an alignment of nucleotide sequences from a variety of expressed sequence tags as well as a consensus nucleotide sequence derived therefrom designated "consen01", and also referred to herein as DNA36459.

Figure 4A to 4B show a double-stranded representation of the DNA36459 sequence (SEQ ID NO:4) shown in Figures 3A to 3H as well as the locations (as underlined) of oligonucleotide primers based upon the DNA36459 consensus sequence, wherein the oligonucleotide primers are herein designated "36459.f1", "36459.r1" and "36459.p1". The deduced amino acid sequence encoded by the DNA36459 sequence is shown below the

nucleotide sequence.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

I. Definitions

The terms "PRO874 polypeptide" and "PRO874" when used herein encompass native sequence PRO874 and PRO874 polypeptide variants (which are further defined herein). The PRO874 polypeptides may be isolated from a variety of sources, such as from human tissue types or from another source, or prepared by recombinant or synthetic methods.

A "native sequence PRO874 polypeptide" comprises a polypeptide having the same amino acid sequence as a PRO874 polypeptide derived from nature. Such native sequence PRO874 polypeptide can be isolated from nature or can be produced by recombinant or synthetic means. The term "native sequence PRO874 polypeptide" specifically encompasses naturally-occurring truncated or secreted forms of a PRO874 polypeptide (*e.g.*, soluble forms containing for instance, an extracellular domain sequence), naturally-occurring variant forms (*e.g.*, alternatively spliced forms) and naturally-occurring allelic variants of a PRO874 polypeptide. In one embodiment of the invention, the native sequence PRO874 polypeptide comprises amino acids 1 to 321 of Figure 2 (SEQ ID NO:3). Optionally, the PRO874 polypeptide is obtained or obtainable by expressing the polypeptide encoded by the cDNA insert of the vector DNA40621-1440 deposited on June 2, 1998, as ATCC ________________.

The "PRO874 extracellular domain" or "PRO874 ECD" refers to a form of the PRO874 polypeptide which is essentially free of the transmembrane and cytoplasmic domains of the PRO874 polypeptide. Optionally the PRO874 ECD will comprise amino acids 275 to 321 of Figure 2 (SEQ ID NO:3). Ordinarily, PRO874 ECD will have less than 1% of such transmembrane and/or cytoplasmic domains and preferably, will have less than 0.5% of such domains. It will be understood that any transmembrane domain identified for the PRO874 polypeptide of the present invention is identified pursuant to criteria routinely employed in the art for identifying that type of hydrophobic domain. The exact boundaries of a transmembrane domain may vary but most likely by no more than about 5 amino acids at either end of the domain as initially identified. Accordingly, the PRO874 polypeptide ECD may optionally comprise amino acids X to 321 of Figure 2 (SEQ ID NO:3), wherein X is any one of amino acid residues 270 to 279 of Figure 2 (SEQ ID NO:3).

"PRO874 variant" means an active PRO874 polypeptide as defined below having at least about 80% amino acid sequence identity with the PRO874 polypeptide having the deduced amino acid sequence shown in Figure 2 (SEQ ID NO:3). Such PRO874 polypeptide variants include, for instance, PRO874 polypeptides wherein one or more amino acid residues are added, or deleted, at the N- or C-terminus of the sequence of Figure 2 (SEQ ID NO:3). Ordinarily, a PRO874 polypeptide variant will have at least about 80% amino acid sequence identity, preferably at least about 85% amino acid sequence identity, more preferably at least about 90% amino acid sequence identity and even more preferably at least about 95% amino acid sequence identity with the amino acid sequence of Figure 2 (SEQ ID NO:3). The term variant does not encompass the known peptides which have identity with the PRO874 described herein.

"Percent (%) amino acid sequence identity" or "percent (%) nucleic acid sequence identity" when referring to the PRO874 amino acid sequences or DNA40621-1440nucleic acid sequences identified herein, respectively, is defined as the percentage amino acid residues or nucleic acid residues in a candidate sequence that are identical with the amino acid residues in a PRO874 polypeptide sequence, or the nucleic acid residues in a DNA40621-1440 nucleic acid sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity. The % identity values used herein were generated using the WU-BLAST-2 computer program (Altschul et al., Methods in Enzymology 266:460-480 (1996); http://blast.wustl/edu/blast/README.html). Most of the WU-BLAST-2 search parameters were set to the default values. The adjustable parameters were set with the following values: overlap span = 1, overlap fraction = 0.125, word threshold (T) = 11, and scoring matrix = BLOSUM62. The HSP S and HSP S2 parameters, which are dynamic values used by BLAST-2, are established by the program itself depending upon the composition of the sequence of interest and composition of the database against which the sequence is being searched. However, the values may be adjusted to increase sensitivity. A % sequence identity value is determined by the fraction of matching identical residues divided by the total number of residues of the "longer" sequence. The "longer" sequence is the one having the most actual residues in the aligned region (gaps introduced by WU-BLAST-2 to maximize the alignment score are ignored).

Percent amino acid or nucleic acid sequence identity can also be determined in other

ways that are within the skill in the art, for instance, using other publicly available computer software such as ALIGN or Megalign (DNASTAR) software. Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the full length of the sequences being compared.

The term "epitope tagged" where used herein refers to a chimeric polypeptide comprising a PRO874 polypeptide, or domain sequence thereof, fused to a "tag polypeptide". The tag polypeptide has enough residues to provide an epitope against which an antibody may be made, or which can be identified by some other agent, yet is short enough such that it does not interfere with the activity of the PRO874 polypeptide. The tag polypeptide preferably is also fairly unique so that the antibody does not substantially cross-react with other epitopes. Suitable tag polypeptides generally have at least six amino acid residues and usually between about 8 to about 50 amino acid residues (preferably, between about 10 to about 20 residues).

"Isolated," when used to describe the various polypeptides disclosed herein, means polypeptide that has been identified and separated and/or recovered from a component of its natural environment. Contaminant components of its natural environment are materials that would typically interfere with diagnostic or therapeutic uses for the polypeptide, and may include enzymes, hormones, and other proteinaceous or non-proteinaceous solutes. In preferred embodiments, the polypeptide will be purified (1) to a degree sufficient to obtain at least 15 residues of N-terminal or internal amino acid sequence by use of a spinning cup sequenator, or (2) to homogeneity by SDS-PAGE under non-reducing or reducing conditions using Coomassie blue or, preferably, silver stain. Isolated polypeptide includes polypeptide *in situ* within recombinant cells, since at least one component of the PRO874 polypeptide natural environment will not be present. Ordinarily, however, isolated polypeptide will be prepared by at least one purification step.

An "isolated" PRO874 polypeptide-encoding nucleic acid molecule is a nucleic acid molecule that is identified and separated from at least one contaminant nucleic acid molecule with which it is ordinarily associated in the natural source of the PRO874 polypeptide-encoding nucleic acid. An isolated PRO874 polypeptide-encoding nucleic acid molecule is other than in the form or setting in which it is found in nature. Isolated PRO874 polypeptide-encoding nucleic acid molecules therefore are distinguished from the PRO874 polypeptide-encoding nucleic acid molecule as it exists in natural cells. However, an isolated PRO874 polypeptide-encoding nucleic acid molecule includes PRO874 polypeptide-encoding nucleic

acid molecules contained in cells that ordinarily express PRO874 polypeptide where, for example, the nucleic acid molecule is in a chromosomal location different from that of natural cells.

The term "control sequences" refers to DNA sequences necessary for the expression of an operably linked coding sequence in a particular host organism. The control sequences that are suitable for prokaryotes, for example, include a promoter, optionally an operator sequence, and a ribosome binding site. Eukaryotic cells are known to utilize promoters, polyadenylation signals, and enhancers.

Nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For example, DNA for a presequence or secretory leader is operably linked to DNA for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, "operably linked" means that the DNA sequences being linked are contiguous, and, in the case of a secretory leader, contiguous and in reading phase. However, enhancers do not have to be contiguous. Linking is accomplished by ligation at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adaptors or linkers are used in accordance with conventional practice.

The term "antibody" is used in the broadest sense and specifically covers single anti-PRO874 polypeptide monoclonal antibodies (including agonist, antagonist, and neutralizing antibodies) and anti-PRO874 antibody compositions with polyepitopic specificity. The term "monoclonal antibody" as used herein refers to an antibody obtained from a population of substantially homogeneous antibodies, *i.e.*, the individual antibodies comprising the population are identical except for possible naturally-occurring mutations that may be present in minor amounts.

"Active" or "activity" for the purposes herein refers to form(s) of PRO874 which retain the biologic and/or immunologic activities of native or naturally-occurring PRO874 polypeptide.

The terms "treating", "treatment" and "therapy" as used herein refer to curative therapy, prophylactic therapy, and preventative therapy.

The term "mammal" as used herein refers to any mammal classified as a mammal,

including humans, cows, horses, dogs and cats. In a preferred embodiment of the invention, the mammal is a human.

II. Compositions and Methods of the Invention

A. Native PRO874 Polypeptide

The present invention provides newly identified and isolated nucleotide sequences encoding polypeptides referred to in the present application as PRO874. In particular, Applicants have identified and isolated cDNA encoding a PRO874 polypeptide, as disclosed in further detail in the Examples below. The PRO874-encoding clone was isolated from a human fetal lung library. To Applicants present knowledge, the UNQ441 (DNA40621-1440) nucleotide sequence encodes a novel factor. Although, as described below in Example 1, using BLAST and FastA sequence alignment computer programs, some sequence identity with known proteins was revealed.

B. PRO874 Variants

In addition to the native sequence PRO874 polypeptide described herein, it is contemplated that PRO874 variants can be prepared. PRO874 variants can be prepared by introducing appropriate nucleotide changes into the PRO874-encoding DNA, or by synthesis of the desired PRO874 polypeptide. Those skilled in the art will appreciate that amino acid changes may alter post-translational processes of the PRO874 polypeptide, such as changing the number or position of glycosylation sites or altering the membrane anchoring characteristics.

Variations in the native sequence of the PRO874 polypeptide or in various domains of the PRO874 polypeptide described herein, can be made, for example, using any of the techniques and guidelines for conservative and non-conservative mutations set forth, for instance, in U.S. Patent No. 5,364,934. Variations may be a substitution, deletion or insertion of one or more codons encoding the PRO874 polypeptide that results in a change in the amino acid sequence of the PRO874 polypeptide as compared with the native sequence PRO874. Optionally the variation is by substitution of at least one amino acid with any other amino acid in one or more of the domains of the PRO874 polypeptide. Guidance in determining which amino acid residue may be inserted, substituted or deleted without adversely affecting the desired activity may be found by comparing the sequence of the PRO874 polypeptide with that

of homologous known protein molecules and minimizing the number of amino acid sequence changes made in regions of high homology. Amino acid substitutions can be the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, such as the replacement of a leucine with a serine, i.e., conservative amino acid replacements. Insertions or deletions may optionally be in the range of 1 to 5 amino acids. The variation allowed may be determined by systematically making insertions, deletions or substitutions of amino acids in the sequence and testing the resulting variants for activity in any of the *in vitro* assays described in the Examples below.

The variations can be made using methods known in the art such as oligonucleotide-mediated (site-directed) mutagenesis, alanine scanning, and PCR mutagenesis. Site-directed mutagenesis [Carter et al., Nucl. Acids Res., 13:4331 (1986); Zoller et al., Nucl. Acids Res., 10:6487 (1987)], cassette mutagenesis [Wells et al., Gene, 34:315 (1985)], restriction selection mutagenesis [Wells et al., Philos. Trans. R. Soc. London SerA, 317:415 (1986)] or other known techniques can be performed on the cloned DNA to produce the PRO874-encoding variant DNA.

Scanning amino acid analysis can also be employed to identify one or more amino acids along a contiguous sequence. Among the preferred scanning amino acids are relatively small, neutral amino acids. Such amino acids include alanine, glycine, serine, and cysteine. Alanine is typically a preferred scanning amino acid among this group because it eliminates the side-chain beyond the beta-carbon and is less likely to alter the main-chain conformation of the variant. Alanine is also typically preferred because it is the most common amino acid. Further, it is frequently found in both buried and exposed positions [Creighton, The Proteins, (W.H. Freeman & Co., N.Y.); Chothia, J. Mol. Biol., 150:1 (1976)]. If alanine substitution does not yield adequate amounts of variant, an isoteric amino acid can be used.

C. Modifications of PRO874

Covalent modifications of PRO874 polypeptides are included within the scope of this invention. One type of covalent modification includes reacting targeted amino acid residues of a PRO874 polypeptide with an organic derivatizing agent that is capable of reacting with selected side chains or the N- or C- terminal residues of a PRO874 polypeptide. Derivatization with bifunctional agents is useful, for instance, for crosslinking PRO874 to a water-insoluble support matrix or surface for use in the method for purifying anti-PRO874

antibodies, and vice-versa. Commonly used crosslinking agents include, *e.g.*, 1,1-bis(diazoacetyl)-2-phenylethane, glutaraldehyde, N-hydroxysuccinimide esters, for example, esters with 4-azidosalicylic acid, homobifunctional imidoesters, including disuccinimidyl esters such as 3,3'-dithiobis(succinimidylpropionate), bifunctional maleimides such as bis-N-maleimido-1,8-octane and agents such as methyl-3-[(p-azidophenyl)dithio]propioimidate.

Other modifications include deamidation of glutaminyl and asparaginyl residues to the corresponding glutamyl and aspartyl residues, respectively, hydroxylation of proline and lysine, phosphorylation of hydroxyl groups of seryl or threonyl residues, methylation of the α-amino groups of lysine, arginine, and histidine side chains [T.E. Creighton, Proteins: Structure and Molecular Properties, W.H. Freeman & Co., San Francisco, pp. 79-86 (1983)], acetylation of the N-terminal amine, and amidation of any C-terminal carboxyl group.

Another type of covalent modification of the PRO874 polypeptide included within the scope of this invention comprises altering the native glycosylation pattern of the polypeptide. "Altering the native glycosylation pattern" is intended for purposes herein to mean deleting one or more carbohydrate moieties found in native sequence PRO874 polypeptide, and/or adding one or more glycosylation sites that are not present in the native sequence PRO874 polypeptide.

Addition of glycosylation sites to PRO874 polypeptides may be accomplished by altering the amino acid sequence thereof. The alteration may be made, for example, by the addition of, or substitution by, one or more serine or threonine residues to the native sequence PRO874 polypeptide (for O-linked glycosylation sites). The PRO874 amino acid sequence may optionally be altered through changes at the DNA level, particularly by mutating the DNA encoding the PRO874 polypeptide at preselected bases such that codons are generated that will translate into the desired amino acids.

Another means of increasing the number of carbohydrate moieties on the PRO874 polypeptide is by chemical or enzymatic coupling of glycosides to the polypeptide. Such methods are described in the art, e.g., in WO 87/05330 published 11 September 1987, and in Aplin and Wriston, CRC Crit. Rev. Biochem., pp. 259-306 (1981).

Removal of carbohydrate moieties present on the PRO874 polypeptide may be accomplished chemically or enzymatically or by mutational substitution of codons encoding for amino acid residues that serve as targets for glycosylation. Chemical deglycosylation techniques are known in the art and described, for instance, by Hakimuddin, et al., Arch.

Biochem. Biophys., 259:52 (1987) and by Edge et al., Anal. Biochem., 118:131 (1981). Enzymatic cleavage of carbohydrate moieties on polypeptides can be achieved by the use of a variety of endo- and exo-glycosidases as described by Thotakura et al., Meth. Enzymol., 138:350 (1987).

Another type of covalent modification of PRO874 comprises linking the PRO874 polypeptide to one of a variety of nonproteinaceous polymers, *e.g.*, polyethylene glycol, polypropylene glycol, or polyoxyalkylenes, in the manner set forth in U.S. Patent Nos. 4,640,835; 4,496,689; 4,301,144; 4,670,417; 4,791,192 or 4,179,337.

PRO874 polypeptides of the present invention may also be modified in a way to form chimeric molecules comprising a PRO874 polypeptide fused to another, heterologous polypeptide or amino acid sequence. In one embodiment, such a chimeric molecule comprises a fusion of a PRO874 polypeptide with a tag polypeptide which provides an epitope to which an anti-tag antibody can selectively bind. The epitope tag is generally placed at the amino- or carboxyl- terminus of the PRO874 polypeptide. The presence of such epitope-tagged forms of a PRO874 polypeptide can be detected using an antibody against the tag polypeptide. Also, provision of the epitope tag enables the PRO874 polypeptide to be readily purified by affinity purification using an anti-tag antibody or another type of affinity matrix that binds to the epitope tag. In an alternative embodiment, the chimeric molecule may comprise a fusion of a PRO874 polypeptide with an immunoglobulin or a particular region of an immunoglobulin. For a bivalent form of the chimeric molecule, such a fusion could be to the Fc region of an IgG molecule.

Various tag polypeptides and their respective antibodies are well known in the art. Examples include poly-histidine (poly-his) or poly-histidine-glycine (poly-his-gly) tags; the flu HA tag polypeptide and its antibody 12CA5 [Field et al., Mol. Cell. Biol., 8:2159-2165 (1988)]; the c-myc tag and the 8F9, 3C7, 6E10, G4, B7 and 9E10 antibodies thereto [Evan et al., Molecular and Cellular Biology, 5:3610-3616 (1985)]; and the Herpes Simplex virus glycoprotein D (gD) tag and its antibody [Paborsky et al., Protein Engineering, 3(6):547-553 (1990)]. Other tag polypeptides include the Flag-peptide [Hopp et al., BioTechnology, 6:1204-1210 (1988)]; the KT3 epitope peptide [Martin et al., Science, 255:192-194 (1992)]; an α-tubulin epitope peptide [Skinner et al., J. Biol. Chem., 266:15163-15166 (1991)]; and the T7 gene 10 protein peptide tag [Lutz-Freyermuth et al., Proc. Natl. Acad. Sci. USA, 87:6393-6397 (1990)].

The PRO874 polypeptide of the present invention may also be modified in a way to form a chimeric molecule comprising a PRO874 polypeptide fused to a leucine zipper. Various leucine zipper polypeptides have been described in the art. See, e.g., Landschulz et al., Science 240:1759 (1988); WO 94/10308; Hoppe et al., FEBS Letters 344:1991 (1994); Maniatis et al., Nature 341:24 (1989). It is believed that use of a leucine zipper fused to a PRO874 polypeptide may be desirable to assist in dimerizing or trimerizing soluble PRO874 polypeptide in solution. Those skilled in the art will appreciate that the leucine zipper may be fused at either the N- or C-terminal end of the PRO874 molecule.

D. Preparation of PRO874

The description below relates primarily to production of PRO874 by culturing cells transformed or transfected with a vector containing PRO874 polypeptide encoding nucleic acid. It is, of course, contemplated that alternative methods, which are well known in the art, may be employed to prepare PRO874 polypeptides. For instance, the PRO874 sequence, or portions thereof, may be produced by direct peptide synthesis using solid-phase techniques [see, e.g., Stewart et al., Solid-Phase Peptide Synthesis, W.H. Freeman Co., San Francisco, CA (1969); Merrifield, J. Am. Chem. Soc., 85:2149-2154 (1963)]. *In vitro* protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be accomplished, for instance, using an Applied Biosystems Peptide Synthesizer (Foster City, CA) using manufacturer's instructions. Various portions of PRO874 polypeptides may be chemically synthesized separately and combined using chemical or enzymatic methods to produce a PRO874 polypeptide.

1. Isolation of DNA Encoding PRO874

DNA encoding a PRO874 polypeptide may be obtained from a cDNA library prepared from tissue believed to possess the PRO874 mRNA and to express it at a detectable level. Accordingly, human PRO874-encoding DNA can be conveniently obtained from a cDNA library prepared from human tissue, such as described in the Examples. The PRO874-encoding gene may also be obtained from a genomic library or by oligonucleotide synthesis.

Libraries can be screened with probes (such as antibodies to a PRO874 polypeptide or oligonucleotides of at least about 20-80 bases) designed to identify the gene of interest or the protein encoded by it. Screening the cDNA or genomic library with the selected probe may

be conducted using standard procedures, such as described in Sambrook et al., Molecular Cloning: A Laboratory Manual (New York: Cold Spring Harbor Laboratory Press, 1989). An alternative means to isolate the gene encoding PRO874 is to use PCR methodology [Sambrook et al., supra; Dieffenbach et al., PCR Primer:A Laboratory Manual (Cold Spring Harbor Laboratory Press, 1995)].

The Examples below describe techniques for screening a cDNA library. The oligonucleotide sequences selected as probes should be of sufficient length and sufficiently unambiguous that false positives are minimized. The oligonucleotide is preferably labeled such that it can be detected upon hybridization to DNA in the library being screened. Methods of labeling are well known in the art, and include the use of radiolabels like ^{32}P-labeled ATP, biotinylation or enzyme labeling. Hybridization conditions, including moderate stringency and high stringency, are provided in Sambrook et al., supra.

Sequences identified in such library screening methods can be compared and aligned to other known sequences deposited and available in public databases such as GenBank or other private sequence databases. Sequence identity (at either the amino acid or nucleotide level) within defined regions of the molecule or across the full-length sequence can be determined through sequence alignment using computer software programs such as ALIGN, DNAstar, and INHERIT which employ various algorithms to measure homology.

Nucleic acid having protein coding sequence may be obtained by screening selected cDNA or genomic libraries using the deduced amino acid sequence disclosed herein for the first time, and, if necessary, using conventional primer extension procedures as described in Sambrook et al., supra, to detect precursors and processing intermediates of mRNA that may not have been reverse-transcribed into cDNA.

2. Selection and Transformation of Host Cells

Host cells are transfected or transformed with expression or cloning vectors described herein for PRO874 polypeptide production and cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences. The culture conditions, such as media, temperature, pH and the like, can be selected by the skilled artisan without undue experimentation. In general, principles, protocols, and practical techniques for maximizing the productivity of cell cultures can be found in Mammalian Cell Biotechnology: a Practical Approach, M. Butler,

ed. (IRL Press, 1991) and Sambrook et al., supra.

Methods of transfection are known to the ordinarily skilled artisan, for example, $CaPO_4$ and electroporation. Depending on the host cell used, transformation is performed using standard techniques appropriate to such cells. The calcium treatment employing calcium chloride, as described in Sambrook et al., supra, or electroporation is generally used for prokaryotes or other cells that contain substantial cell-wall barriers. Infection with *Agrobacterium tumefaciens* is used for transformation of certain plant cells, as described by Shaw et al., Gene, 23:315 (1983) and WO 89/05859 published June 29, 1989. For mammalian cells without such cell walls, the calcium phosphate precipitation method of Graham and van der Eb, Virology, 52:456-457 (1978) can be employed. General aspects of mammalian cell host system transformations have been described in U.S. Patent No. 4,399,216. Transformations into yeast are typically carried out according to the method of Van Solingen et al., J. Bact., 130:946 (1977) and Hsiao et al., Proc. Natl. Acad. Sci. (USA), 76:3829 (1979). However, other methods for introducing DNA into cells, such as by nuclear microinjection, electroporation, bacterial protoplast fusion with intact cells, or polycations, *e.g.*, polybrene, polyornithine, may also be used. For various techniques for transforming mammalian cells, see Keown et al., Methods in Enzymology, 185:527-537 (1990) and Mansour et al., Nature, 336:348-352 (1988).

Suitable host cells for cloning or expressing the DNA in the vectors herein include prokaryote, yeast, or higher eukaryote cells. Suitable prokaryotes include but are not limited to eubacteria, such as Gram-negative or Gram-positive organisms, for example, Enterobacteriaceae such as *E. coli*. Various *E. coli* strains are publicly available, such as *E. coli* K12 strain MM294 (ATCC 31,446); *E. coli* X1776 (ATCC 31,537); *E. coli* strain W3110 (ATCC 27,325) and K5 772 (ATCC 53,635).

In addition to prokaryotes, eukaryotic microbes such as filamentous fungi or yeast are suitable cloning or expression hosts for PRO874-encoding vectors. *Saccharomyces cerevisiae* is a commonly used lower eukaryotic host microorganism.

Suitable host cells for the expression of glycosylated PRO874 are derived from multicellular organisms. Examples of invertebrate cells include insect cells such as Drosophila S2 and Spodoptera Sf9, as well as plant cells. Examples of useful mammalian host cell lines include Chinese hamster ovary (CHO) and COS cells. More specific examples include monkey kidney CV1 line transformed by SV40 (COS-7, ATCC CRL 1651); human embryonic

kidney line (293 or 293 cells subcloned for growth in suspension culture, Graham et al., J. Gen Virol., 36:59 (1977)); Chinese hamster ovary cells/-DHFR (CHO, Urlaub and Chasin, Proc. Natl. Acad. Sci. USA, 77:4216 (1980)); mouse sertoli cells (TM4, Mather, Biol. Reprod., 23:243-251 (1980)); human lung cells (W138, ATCC CCL 75); human liver cells (Hep G2, HB 8065); and mouse mammary tumor (MMT 060562, ATCC CCL51). The selection of the appropriate host cell is deemed to be within the skill in the art.

3. Selection and Use of a Replicable Vector

The nucleic acid (*e.g.*, cDNA or genomic DNA) encoding the desired PRO874 polypeptide may be inserted into a replicable vector for cloning (amplification of the DNA) or for expression. Various vectors are publicly available. The vector may, for example, be in the form of a plasmid, cosmid, viral particle, or phage. The appropriate nucleic acid sequence may be inserted into the vector by a variety of procedures. In general, DNA is inserted into an appropriate restriction endonuclease site(s) using techniques known in the art. Vector components generally include, but are not limited to, one or more of a signal sequence, an origin of replication, one or more marker genes, an enhancer element, a promoter, and a transcription termination sequence. Construction of suitable vectors containing one or more of these components employs standard ligation techniques which are known to the skilled artisan.

The desired PRO874 polypeptide may be produced recombinantly not only directly, but also as a fusion polypeptide with a heterologous polypeptide, which may be a signal sequence or other polypeptide having a specific cleavage site at the N-terminus of the mature protein or polypeptide. In general, the signal sequence may be a component of the vector, or it may be a part of the PRO874-encoding DNA that is inserted into the vector. The signal sequence may be a prokaryotic signal sequence selected, for example, from the group of the alkaline phosphatase, penicillinase, lpp, or heat-stable enterotoxin II leaders. For yeast secretion the signal sequence may be, *e.g.*, the yeast invertase leader, alpha factor leader (including *Saccharomyces* and *Kluyveromyces* α-factor leaders, the latter described in U.S. Patent No. 5,010,182), or acid phosphatase leader, the *C. albicans* glucoamylase leader (EP 362,179 published 4 April 1990), or the signal described in WO 90/13646 published 15 November 1990. In mammalian cell expression, mammalian signal sequences may be used to direct secretion of the protein, such as signal sequences from secreted polypeptides of the

same or related species, as well as viral secretory leaders.

Both expression and cloning vectors contain a nucleic acid sequence that enables the vector to replicate in one or more selected host cells. Such sequences are well known for a variety of bacteria, yeast, and viruses. The origin of replication from the plasmid pBR322 is suitable for most Gram-negative bacteria, the 2μ plasmid origin is suitable for yeast, and various viral origins (SV40, polyoma, adenovirus, VSV or BPV) are useful for cloning vectors in mammalian cells.

Expression and cloning vectors will typically contain a selection gene, also termed a selectable marker. Typical selection genes encode proteins that (a) confer resistance to antibiotics or other toxins, *e.g.*, ampicillin, neomycin, methotrexate, or tetracycline, (b) complement auxotrophic deficiencies, or (c) supply critical nutrients not available from complex media, *e.g.*, the gene encoding D-alanine racemase for *Bacilli*.

An example of suitable selectable markers for mammalian cells are those that enable the identification of cells competent to take up the PRO874-encoding nucleic acid, such as DHFR or thymidine kinase. An appropriate host cell when wild-type DHFR is employed is the CHO cell line deficient in DHFR activity, prepared and propagated as described by Urlaub et al., Proc. Natl. Acad. Sci. USA, 77:4216 (1980). A suitable selection gene for use in yeast is the *trp*1 gene present in the yeast plasmid YRp7 [Stinchcomb et al., Nature, 282:39 (1979); Kingsman et al., Gene, 7:141 (1979); Tschemper et al., Gene, 10:157 (1980)]. The *trp*1 gene provides a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan, for example, ATCC No. 44076 or PEP4-1 [Jones, Genetics, 85:12 (1977)].

Expression and cloning vectors usually contain a promoter operably linked to the PRO874-encoding nucleic acid sequence to direct mRNA synthesis. Promoters recognized by a variety of potential host cells are well known. Promoters suitable for use with prokaryotic hosts include the β-lactamase and lactose promoter systems [Chang et al., Nature, 275:615 (1978); Goeddel et al., Nature, 281:544 (1979)], alkaline phosphatase, a tryptophan (trp) promoter system [Goeddel, Nucleic Acids Res., 8:4057 (1980); EP 36,776], and hybrid promoters such as the tac promoter [deBoer et al., Proc. Natl. Acad. Sci. USA, 80:21-25 (1983)]. Promoters for use in bacterial systems also will contain a Shine-Dalgarno (S.D.) sequence operably linked to the DNA encoding the PRO874 polypeptide.

Examples of suitable promoting sequences for use with yeast hosts include the promoters for 3-phosphoglycerate kinase [Hitzeman et al., J. Biol. Chem., 255:2073 (1980)]

or other glycolytic enzymes [Hess et al., J. Adv. Enzyme Reg., 7:149 (1968); Holland, Biochemistry, 17:4900 (1978)], such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase.

Other yeast promoters, which are inducible promoters having the additional advantage of transcription controlled by growth conditions, are the promoter regions for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, metallothionein, glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization. Suitable vectors and promoters for use in yeast expression are further described in EP 73,657.

PRO874 transcription from vectors in mammalian host cells is controlled, for example, by promoters obtained from the genomes of viruses such as polyoma virus, fowlpox virus (UK 2,211,504 published 5 July 1989), adenovirus (such as Adenovirus 2), bovine papilloma virus, avian sarcoma virus, cytomegalovirus, a retrovirus, hepatitis-B virus and Simian Virus 40 (SV40), from heterologous mammalian promoters, *e.g.*, the actin promoter or an immunoglobulin promoter, and from heat-shock promoters, provided such promoters are compatible with the host cell systems.

Transcription of a DNA encoding a PRO874 polypeptide by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp, that act on a promoter to increase its transcription. Many enhancer sequences are now known from mammalian genes (globin, elastase, albumin, α-fetoprotein, and insulin). Typically, however, one will use an enhancer from a eukaryotic cell virus. Examples include the SV40 enhancer on the late side of the replication origin (bp 100-270), the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers. The enhancer may be spliced into the vector at a position 5' or 3' to the PRO874 coding sequence, but is preferably located at a site 5' from the promoter.

Expression vectors used in eukaryotic host cells (yeast, fungi, insect, plant, animal, human, or nucleated cells from other multicellular organisms) will also contain sequences necessary for the termination of transcription and for stabilizing the mRNA. Such sequences are commonly available from the 5' and, occasionally 3', untranslated regions of eukaryotic

or viral DNAs or cDNAs. These regions contain nucleotide segments transcribed as polyadenylated fragments in the untranslated portion of the mRNA encoding PRO874.

Still other methods, vectors, and host cells suitable for adaptation to the synthesis of PRO874 polypeptides in recombinant vertebrate cell culture are described in Gething et al., Nature, 293:620-625 (1981); Mantei et al., Nature, 281:40-46 (1979); EP 117,060; and EP 117,058.

4. Detecting Gene Amplification/Expression

Gene amplification and/or expression may be measured in a sample directly, for example, by conventional Southern blotting, Northern blotting to quantitate the transcription of mRNA [Thomas, Proc. Natl. Acad. Sci. USA, 77:5201-5205 (1980)], dot blotting (DNA analysis), or *in situ* hybridization, using an appropriately labeled probe, based on the sequences provided herein. Alternatively, antibodies may be employed that can recognize specific duplexes, including DNA duplexes, RNA duplexes, and DNA-RNA hybrid duplexes or DNA-protein duplexes. The antibodies in turn may be labeled and the assay may be carried out where the duplex is bound to a surface, so that upon the formation of duplex on the surface, the presence of antibody bound to the duplex can be detected.

Gene expression, alternatively, may be measured by immunological methods, such as immunohistochemical staining of cells or tissue sections and assay of cell culture or body fluids, to quantitate directly the expression of gene product. Antibodies useful for immunohistochemical staining and/or assay of sample fluids may be either monoclonal or polyclonal, and may be prepared in any mammal. Conveniently, the antibodies may be prepared against a native sequence PRO874 polypeptide or against a synthetic peptide based on the DNA sequences provided herein or against exogenous sequence fused to PRO874-encoding DNA and encoding a specific antibody epitope.

5. Purification of Polypeptide

Forms of PRO874 may be recovered from culture medium or from host cell lysates. If membrane-bound, it can be released from the membrane using a suitable detergent solution (*e.g.* Triton-X 100) or by enzymatic cleavage. Cells employed in expression of PRO874 polypeptides can be disrupted by various physical or chemical means, such as freeze-thaw cycling, sonication, mechanical disruption, or cell lysing agents.

It may be desired to purify PRO874 from recombinant cell proteins or polypeptides. The following procedures are exemplary of suitable purification procedures: by fractionation on an ion-exchange column; ethanol precipitation; reverse phase HPLC; chromatography on silica or on a cation-exchange resin such as DEAE; chromatofocusing; SDS-PAGE; ammonium sulfate precipitation; gel filtration using, for example, Sephadex G-75; protein A Sepharose columns to remove contaminants such as IgG; and metal chelating columns to bind epitope-tagged forms of the PRO874 polypeptide. Various methods of protein purification may be employed and such methods are known in the art and described for example in Deutscher, Methods in Enzymology, 182 (1990); Scopes, Protein Purification:Principles and Practice, Springer-Verlag, New York (1982). The purification step(s) selected will depend, for example, on the nature of the production process used and the particular PRO874 polypeptide produced.

E. Uses for PRO874

Nucleotide sequences (or their complement) encoding PRO874 polypeptides have various applications in the art of molecular biology, including uses as hybridization probes, in chromosome and gene mapping and in the generation of anti-sense RNA and DNA. PRO874-encoding nucleic acid will also be useful for the preparation of PRO874 polypeptides by the recombinant techniques described herein.

The DNA40621-1440 nucleotide sequence (SEQ ID NO:1) or the nucleotide sequence encoding PRO874 (SEQ ID NO:2), or portions thereof, may be used as hybridization probes for a cDNA library to isolate the full-length PRO874 gene or to isolate still other genes (for instance, those encoding naturally-occurring variants of PRO874 or PRO874 from other species) which have a desired sequence identity to the PRO874 nucleotide sequence disclosed in Figure 1 (SEQ ID NO:1). Optionally, the length of the probes will be about 20 to about 50 bases. The hybridization probes may be derived from the UNQ441 (DNA40621-1440) nucleotide sequence of SEQ ID NO:1 as shown in Figure 1 or from genomic sequences including promoters, enhancer elements and introns of native sequence PRO874-encoding DNA. By way of example, a screening method will comprise isolating the coding region of the PRO874 gene using the known DNA sequence to synthesize a selected probe of about 40 bases. Hybridization probes may be labeled by a variety of labels, including radionucleotides such as ^{32}P or ^{35}S, or enzymatic labels such as alkaline phosphatase coupled to the probe via

avidin/biotin coupling systems. Labeled probes having a sequence complementary to that of the PRO874 gene of the present invention can be used to screen libraries of human cDNA, genomic DNA or mRNA to determine which members of such libraries the probe hybridizes to. Hybridization techniques are described in further detail in the Examples below.

The probes may also be employed in PCR techniques to generate a pool of sequences for identification of closely related PRO874 sequences.

Nucleotide sequences encoding a PRO874 polypeptide can also be used to construct hybridization probes for mapping the gene which encodes that PRO874 polypeptide and for the genetic analysis of individuals with genetic disorders. The nucleotide sequences provided herein may be mapped to a chromosome and specific regions of a chromosome using known techniques, such as *in situ* hybridization, linkage analysis against known chromosomal markers, and hybridization screening with libraries.

When the coding sequences for PRO874 encode a protein which binds to another protein (example, where the PRO874 polypeptide functions as a receptor), the PRO874 polypeptide can be used in assays to identify the other proteins or molecules involved in the binding interaction. By such methods, inhibitors of the receptor/ligand binding interaction can be identified. Proteins involved in such binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction. Also, the receptor PRO874 polypeptide can be used to isolate correlative ligand(s). Screening assays can be designed to find lead compounds that mimic the biological activity of a native PRO874 or a receptor for PRO874. Such screening assays will include assays amenable to high-throughput screening of chemical libraries, making them particularly suitable for identifying small molecule drug candidates. Small molecules contemplated include synthetic organic or inorganic compounds. The assays can be performed in a variety of formats, including protein-protein binding assays, biochemical screening assays, immunoassays and cell based assays, which are well characterized in the art.

Nucleic acids which encode PRO874 polypeptide or any of its modified forms can also be used to generate either transgenic animals or "knock out" animals which, in turn, are useful in the development and screening of therapeutically useful reagents. A transgenic animal (e.g., a mouse or rat) is an animal having cells that contain a transgene, which transgene was introduced into the animal or an ancestor of the animal at a prenatal, e.g., an embryonic stage. A transgene is a DNA which is integrated into the genome of a cell from which a transgenic

animal develops. In one embodiment, cDNA encoding PRO874 polypeptide can be used to clone genomic DNA encoding PRO874 in accordance with established techniques and the genomic sequences used to generate transgenic animals that contain cells which express DNA encoding PRO874. Methods for generating transgenic animals, particularly animals such as mice or rats, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009. Typically, particular cells would be targeted for PRO874 transgene incorporation with tissue-specific enhancers. Transgenic animals that include a copy of a transgene encoding PRO874 introduced into the germ line of the animal at an embryonic stage can be used to examine the effect of increased expression of DNA encoding PRO874. Such animals can be used as tester animals for reagents thought to confer protection from, for example, pathological conditions associated with its overexpression. In accordance with this facet of the invention, an animal is treated with the reagent and a reduced incidence of the pathological condition, compared to untreated animals bearing the transgene, would indicate a potential therapeutic intervention for the pathological condition.

Alternatively, non-human homologues of PRO874 can be used to construct a PRO874 "knock out" animal which has a defective or altered gene encoding PRO874 as a result of homologous recombination between the endogenous gene encoding PRO874 and altered genomic DNA encoding PRO874 introduced into an embryonic cell of the animal. For example, cDNA encoding PRO874 can be used to clone genomic DNA encoding PRO874 in accordance with established techniques. A portion of the genomic DNA encoding PRO874 can be deleted or replaced with another gene, such as a gene encoding a selectable marker which can be used to monitor integration. Typically, several kilobases of unaltered flanking DNA (both at the 5' and 3' ends) are included in the vector [see e.g., Thomas and Capecchi, Cell, 51:503 (1987) for a description of homologous recombination vectors]. The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced DNA has homologously recombined with the endogenous DNA are selected [see e.g., Li et al., Cell, 69:915 (1992)]. The selected cells are then injected into a blastocyst of an animal (e.g., a mouse or rat) to form aggregation chimeras [see e.g., Bradley, in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E. J. Robertson, ed. (IRL, Oxford, 1987), pp. 113-152]. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term to create a "knock out" animal. Progeny harboring the homologously recombined DNA in their germ cells can be

identified by standard techniques and used to breed animals in which all cells of the animal contain the homologously recombined DNA. Knockout animals can be characterized for instance, for their ability to defend against certain pathological conditions and for their development of pathological conditions due to absence of the PRO874 polypeptide.

PRO874 polypeptides of the present invention may be employed both *in vivo* for therapeutic purposes and *in vitro*. Those of ordinary skill in the art will well know how to employ the PRO874 polypeptides of the present invention for such purposes.

F. Anti-PRO874 Antibodies

The present invention further provides anti-PRO874 polypeptide antibodies. Exemplary antibodies include polyclonal, monoclonal, humanized, bispecific, and heteroconjugate antibodies.

1. Polyclonal Antibodies

The anti-PRO874 antibodies of the present invention may comprise polyclonal antibodies. Methods of preparing polyclonal antibodies are known to the skilled artisan. Polyclonal antibodies can be raised in a mammal, for example, by one or more injections of an immunizing agent and, if desired, an adjuvant. Typically, the immunizing agent and/or adjuvant will be injected in the mammal by multiple subcutaneous or intraperitoneal injections. The immunizing agent may include the PRO874 polypeptide or a fusion protein thereof. It may be useful to conjugate the immunizing agent to a protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. Examples of adjuvants which may be employed include Freund's complete adjuvant and MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate). The immunization protocol may be selected by one skilled in the art without undue experimentation.

2. Monoclonal Antibodies

The anti-PRO874 antibodies may, alternatively, be monoclonal antibodies. Monoclonal antibodies may be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other

appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes may be immunized *in vitro*.

The immunizing agent will typically include the PRO874 polypeptide or a fusion protein thereof. Generally, either peripheral blood lymphocytes ("PBLs") are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell [Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103]. Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells may be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Rockville, Maryland. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies [Kozbor, J. Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63].

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against a PRO874 polypeptide. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an *in vitro* binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by

the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107:220 (1980).

After the desired hybridoma cells are identified, the clones may be subcloned by limiting dilution procedures and grown by standard methods [Goding, supra]. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells may be grown *in vivo* as ascites in a mammal.

The monoclonal antibodies secreted by the subclones may be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies may also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA may be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also may be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences [U.S. Patent No. 4,816,567; Morrison *et al.*, Proc. Natl. Acad. Sci. USA 81, 6851-6855 [1984]] or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

The antibodies may be monovalent antibodies. Methods for preparing monovalent antibodies are well known in the art. For example, one method involves recombinant expression of immunoglobulin light chain and modified heavy chain. The heavy chain is truncated generally at any point in the Fc region so as to prevent heavy chain crosslinking. Alternatively, the relevant cysteine residues are substituted with another amino acid residue

or are deleted so as to prevent crosslinking.

In vitro methods are also suitable for preparing monovalent antibodies. Digestion of antibodies to produce fragments thereof, particularly, Fab fragments, can be accomplished using routine techniques known in the art.

3. Humanized Antibodies

The anti-PRO874 antibodies of the invention may further comprise humanized antibodies or human antibodies. Humanized forms of non-human (e.g., murine) antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', $F(ab')_2$ or other antigen-binding subsequences of antibodies) which contain minimal sequence derived from non-human immunoglobulin. Humanized antibodies include human immunoglobulins (recipient antibody) in which residues from a complementary determining region (CDR) of the recipient are replaced by residues from a CDR of a non-human species (donor antibody) such as mouse, rat or rabbit having the desired specificity, affinity and capacity. In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies may also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the FR regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin [Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-329 (1988); and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)].

Methods for humanizing non-human antibodies are well known in the art. Generally, a humanized antibody has one or more amino acid residues introduced into it from a source which is non-human. These non-human amino acid residues are often referred to as "import" residues, which are typically taken from an "import" variable domain. Humanization can be essentially performed following the method of Winter and co-workers [Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-327 (1988); Verhoeyen et al., Science, 239:1534-1536 (1988)], by substituting rodent CDRs or CDR sequences for the

corresponding sequences of a human antibody. Accordingly, such "humanized" antibodies are chimeric antibodies (U.S. Patent No. 4,816,567), wherein substantially less than an intact human variable domain has been substituted by the corresponding sequence from a non-human species. In practice, humanized antibodies are typically human antibodies in which some CDR residues and possibly some FR residues are substituted by residues from analogous sites in rodent antibodies.

Human antibodies can also be produced using various techniques known in the art, including phage display libraries [Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)]. The techniques of Cole et al. and Boerner et al. are also available for the preparation of human monoclonal antibodies (Cole et al., Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, p. 77 (1985) and Boerner et al., J. Immunol., 147(1):86-95 (1991)].

4. Bispecific Antibodies

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for a PRO874 polypeptide, the other one is for any other antigen, and preferably for a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities [Milstein and Cuello, Nature, 305:537-539 (1983)]. Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker et al., EMBO J., 10:3655-3659 (1991).

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the

fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., Methods in Enzymology, 121:210 (1986).

5. Heteroconjugate Antibodies

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells [U.S. Patent No. 4,676,980], and for treatment of HIV infection [WO 91/00360; WO 92/200373; EP 03089]. It is contemplated that the antibodies may be prepared *in vitro* using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins may be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

G. Uses for anti-PRO874 Antibodies

The anti-PRO874 antibodies of the present invention have various utilities. For example, anti-PRO874 antibodies may be used in diagnostic assays for PRO874 polypeptides, *e.g.*, detecting expression in specific cells, tissues, or serum. Various diagnostic assay techniques known in the art may be used, such as competitive binding assays, direct or indirect sandwich assays and immunoprecipitation assays conducted in either heterogeneous or homogeneous phases [Zola, Monoclonal Antibodies: A Manual of Techniques, CRC Press, Inc. (1987) pp. 147-158]. The antibodies used in the diagnostic assays can be labeled with a detectable moiety. The detectable moiety should be capable of producing, either directly or indirectly, a detectable signal. For example, the detectable moiety may be a radioisotope, such as ^{3}H, ^{14}C, ^{32}P, ^{35}S, or ^{125}I, a fluorescent or chemiluminescent compound, such as fluorescein isothiocyanate, rhodamine, or luciferin, or an enzyme, such as alkaline phosphatase, beta-galactosidase or horseradish peroxidase. Any method known in the art for conjugating the antibody to the detectable moiety may be employed, including those methods described by Hunter et al., Nature, 144:945 (1962); David et al., Biochemistry, 13:1014 (1974); Pain et

al., J. Immunol. Meth., 40:219 (1981); and Nygren, J. Histochem. and Cytochem., 30:407 (1982).

Anti-PRO874 antibodies also are useful for the affinity purification of PRO874 polypeptides from recombinant cell culture or natural sources. In this process, the antibodies against a PRO874 polypeptide are immobilized on a suitable support, such a Sephadex resin or filter paper, using methods well known in the art. The immobilized antibody then is contacted with a sample containing the PRO874 polypeptide to be purified, and thereafter the support is washed with a suitable solvent that will remove substantially all the material in the sample except the PRO874 polypeptide, which is bound to the immobilized antibody. Finally, the support is washed with another suitable solvent that will release the PRO874 polypeptide from the antibody.

The following examples are offered for illustrative purposes only, and are not intended to limit the scope of the present invention in any way.

All patent and literature references cited in the present specification are hereby incorporated by reference in their entirety.

EXAMPLES

Commercially available reagents referred to in the examples were used according to manufacturer's instructions unless otherwise indicated. The source of those cells identified in the following examples, and throughout the specification, by ATCC accession numbers is the American Type Culture Collection, Rockville, Maryland.

EXAMPLE 1: Isolation of cDNA Clones Encoding Human PRO874

The extracellular domain (ECD) sequences (including the secretion signal, if any) of from about 950 known secreted proteins from the Swiss-Prot public protein database were used to search expressed sequence tag (EST) databases. The EST databases included public EST databases (e.g., GenBank) and a proprietary EST DNA database (LIFESEQ™, Incyte Pharmaceuticals, Palo Alto, CA). The search was performed using the computer program BLAST or BLAST2 (Altshul et al., Methods in Enzymology 266:460-480 (1996)) as a comparison of the ECD protein sequences to a 6 frame translation of the EST sequence. Those comparisons resulting in a BLAST score of 70 (or in some cases 90) or greater that did not encode known proteins were clustered and assembled into consensus DNA sequences with

the program "phrap" (Phil Green, University of Washington, Seattle, Washington; http://bozeman.mbt.washington.edu/phrap.docs/phrap.html).

A consensus DNA sequence was identified using the above described analysis as encoding a potential secreted protein. The consensus sequence, designated herein and in Figure 3 as DNA36459.init sequence, was then extended using repeated cycles of BLAST and phrap to extend the sequence as far as possible using the sources of EST sequences discussed above. The extended assembly sequence is referred to in the alignment figure shown in Figure 3, as "consen01", and is also referred to herein as "DNA36459". The double-stranded sequence of DNA36459 (SEQ ID NO:4) is shown in Figure 4.

Based on the DNA36459 consensus sequence shown in Figure 4, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the coding sequence for PRO874. Forward and reverse PCR primers generally range from 20 to 30 nucleotides and are often designed to give a PCR product of about 100-1000 bp in length. The probe sequences are typically 40-55 bp in length. In some cases, additional oligonucleotides are synthesized when the consensus sequence is greater than about 1-1.5kbp. In order to screen several libraries for a clone, DNA from the libraries was screened by PCR amplification, as per Ausubel et al., Current Protocols in Molecular Biology, with the PCR primer pair. A positive library was then used to isolate clones encoding the gene of interest using the probe oligonucleotide and one of the primer pairs.

PCR primers (forward and reverse) were synthesized:

forward PCR primer (36459.f1) 5'-TCGTGCCCAGGGGCTGATGTGG-3' (SEQ ID NO:5);

and

reverse PCR primer (36459.r1) 5'-GTCTTTACCCAGCCCCGGGATGCG-3' (SEQ ID NO:6).

Additionally, a synthetic oligonucleotide hybridization probe was constructed from the consensus DNA36459 sequence which had the following nucleotide sequence:

hybridization probe (36459.p1)

5'-GGCCTAATCCAACGTTCTGTCTTCAATCTGCAAATCTATGGGGTCCTGGG-3' (SEQ ID NO:7).

In order to screen several libraries for a source of a clone, DNA from the libraries was screened by PCR amplification with the PCR primer pair identified above. A positive library

was then used to isolate clones encoding the PRO874 gene using the probe oligonucleotide and one of the PCR primers.

RNA for construction of the cDNA libraries was isolated from human fetal lung tissue (LIB25). The cDNA libraries used to isolate the cDNA clones were constructed by standard methods using commercially available reagents such as those from Invitrogen, San Diego, CA. The cDNA was primed with oligo dT containing a NotI site, linked with blunt to SalI hemikinased adaptors, cleaved with NotI, sized appropriately by gel electrophoresis, and cloned in a defined orientation into a suitable cloning vector (such as pRKB or pRKD; pRK5B is a precursor of pRK5D that does not contain the SfiI site; see, Holmes et al., Science, 253:1278-1280 (1991)) in the unique XhoI and NotI sites.

DNA sequencing of the clones isolated as described above gave the DNA sequence for PRO874 [herein designated as UNQ441 (DNA40621-1440)] (SEQ ID NO:1) and the derived protein sequence for PRO874.

The entire nucleotide sequence of UNQ441 (DNA40621-1440) is shown in Figure 1 (SEQ ID NO:1). Clone UNQ441 (DNA40621-1440) contains a single open reading frame ending at the stop codon at nucleotide positions 964-966 (Figure 1). The predicted polypeptide encoded by DNA40621-1440 is 321 amino acids long (Figure 2). The PRO874 protein shown in Figure 2 has an estimated molecular weight of about 36,194 daltons and a pI of about 9.85. Analysis of the PRO874 sequence shown in Figure 2 (SEQ ID NO:3) evidenced the presence of the following: a Type II transmembrane domain at about amino acids 57-80; additional transmembrane domains at about amino acids 110-126, 215-231, and 254-274; potential N-glycosylation sites at about amino acids 16-19, 27-30, and 289-292; sequence identity with hypothetical YBR002c family proteins at about amino acids 276-287; and sequence identity with ammonium transporter proteins at about amino acids 204-230.

Clone UNQ441 (DNA40621-1440) was deposited with the ATCC on June 2, 1998, and is assigned ATCC deposit no. ____________________.

Analysis of the amino acid sequence of the PRO874 polypeptide suggests that it is a novel multi-span transmembrane protein. However, an analysis of the Dayhoff database (version 35.45 SwissProt 35) evidenced sequence identity between the PRO874 amino acid sequence and the following Dayhoff sequences: S67049, AF054839_1, S73437, S52460, and HIVU80570_1.

EXAMPLE 2: Use of PRO874-encoding DNA as a hybridization probe

The following method describes use of a nucleotide sequence encoding PRO874 as a hybridization probe.

DNA comprising the coding sequence of PRO874 (as shown in Figure 1, SEQ ID NO:1) or a fragment thereof is employed as a probe to screen for homologous DNAs (such as those encoding naturally-occurring variants of PRO874) in human tissue cDNA libraries or human tissue genomic libraries.

Hybridization and washing of filters containing either library DNAs is performed under the following high stringency conditions. Hybridization of radiolabeled PRO874 polypeptide-derived probe to the filters is performed in a solution of 50% formamide, 5x SSC, 0.1% SDS, 0.1% sodium pyrophosphate, 50 mM sodium phosphate, pH 6.8, 2x Denhardt's solution, and 10% dextran sulfate at 42°C for 20 hours. Washing of the filters is performed in an aqueous solution of 0.1x SSC and 0.1% SDS at 42°C.

DNAs having a desired sequence identity with the DNA encoding the PRO874 polypeptide can then be identified using standard techniques known in the art.

EXAMPLE 3: Expression of PRO874 Polypeptides in *E. coli*

This example illustrates the preparation of unglycosylated forms of PRO874 polypeptides by recombinant expression in *E. coli*.

The DNA sequence encoding the PRO874 polypeptide or a fragment or variant thereof is initially amplified using selected PCR primers. The primers should contain restriction enzyme sites which correspond to the restriction enzyme sites on the selected expression vector. A variety of expression vectors may be employed. An example of a suitable vector is pBR322 (derived from *E. coli*; see Bolivar et al., Gene, 2:95 (1977)) which contains genes for ampicillin and tetracycline resistance. The vector is digested with restriction enzyme and dephosphorylated. The PCR amplified sequences are then ligated into the vector. The vector will preferably include sequences which encode for an antibiotic resistance gene, a trp promoter, a polyhis leader (including the first six STII codons, polyhis sequence, and enterokinase cleavage site), the PRO874 coding region, lambda transcriptional terminator, and an argU gene.

The ligation mixture is then used to transform a selected *E. coli* strain using the methods described in Sambrook et al., supra. Transformants are identified by their ability to

grow on LB plates and antibiotic resistant colonies are then selected. Plasmid DNA can be isolated and confirmed by restriction analysis and DNA sequencing.

Selected clones can be grown overnight in liquid culture medium such as LB broth supplemented with antibiotics. The overnight culture may subsequently be used to inoculate a larger scale culture. The cells are then grown to a desired optical density, during which the expression promoter is turned on.

After culturing the cells for several more hours, the cells can be harvested by centrifugation. The cell pellet obtained by the centrifugation can be solubilized using various agents known in the art, and the solubilized PRO874 polypeptide can then be purified using a metal chelating column under conditions that allow tight binding of the polypeptide.

EXAMPLE 4: Expression of PRO874 Polypeptides in Mammalian Cells

This example illustrates preparation of glycosylated forms of PRO874 polypeptides by recombinant expression in mammalian cells.

The vector, pRK5 (see EP 307,247, published March 15, 1989), is employed as the expression vector. Optionally, the PRO874-encoding DNA is ligated into pRK5 with selected restriction enzymes to allow insertion of the PRO874-encoding DNA using ligation methods such as described in Sambrook et al., supra. The resulting vector is called pRK5-PRO874.

In one embodiment, the selected host cells may be 293 cells. Human 293 cells (ATCC CCL 1573) are grown to confluence in tissue culture plates in medium such as DMEM supplemented with fetal calf serum and optionally, nutrient components and/or antibiotics. About 10 μg pRK5-PRO874 DNA is mixed with about 1 μg DNA encoding the VA RNA gene [Thimmappaya et al., Cell, 31:543 (1982)] and dissolved in 500 μl of 1 mM Tris-HCl, 0.1 mM EDTA, 0.227 M $CaCl_2$. To this mixture is added, dropwise, 500 μl of 50 mM HEPES (pH 7.35), 280 mM NaCl, 1.5 mM $NaPO_4$, and a precipitate is allowed to form for 10 minutes at 25°C. The precipitate is suspended and added to the 293 cells and allowed to settle for about four hours at 37°C. The culture medium is aspirated off and 2 ml of 20% glycerol in PBS is added for 30 seconds. The 293 cells are then washed with serum free medium, fresh medium is added and the cells are incubated for about 5 days.

Approximately 24 hours after the transfections, the culture medium is removed and replaced with culture medium (alone) or culture medium containing 200 μCi/ml ^{35}S-cysteine and 200 μCi/ml ^{35}S-methionine. After a 12 hour incubation, the conditioned medium is

collected, concentrated on a spin filter, and loaded onto a 15% SDS gel. The processed gel may be dried and exposed to film for a selected period of time to reveal the presence of PRO874 polypeptide. The cultures containing transfected cells may undergo further incubation (in serum free medium) and the medium is tested in selected bioassays.

In an alternative technique, PRO874-encoding DNA may be introduced into 293 cells transiently using the dextran sulfate method described by Somparyrac et al., Proc. Natl. Acad. Sci., 12:7575 (1981). 293 cells are grown to maximal density in a spinner flask and 700 μg pRK5-PRO874 DNA is added. The cells are first concentrated from the spinner flask by centrifugation and washed with PBS. The DNA-dextran precipitate is incubated on the cell pellet for four hours. The cells are treated with 20% glycerol for 90 seconds, washed with tissue culture medium, and re-introduced into the spinner flask containing tissue culture medium, 5 μg/ml bovine insulin and 0.1 μg/ml bovine transferrin. After about four days, the conditioned media is centrifuged and filtered to remove cells and debris. The sample containing expressed PRO874 polypeptide can then be concentrated and purified by any selected method, such as dialysis and/or column chromatography.

In another embodiment, PRO874 polypeptide can be expressed in CHO cells. The pRK5-PRO874 vector can be transfected into CHO cells using known reagents such as $CaPO_4$ or DEAE-dextran. As described above, the cell cultures can be incubated, and the medium replaced with culture medium (alone) or medium containing a radiolabel such as ^{35}S-methionine. After determining the presence of PRO874 polypeptide, the culture medium may be replaced with serum free medium. Preferably, the cultures are incubated for about 6 days, and then the conditioned medium is harvested. The medium containing the expressed PRO874 polypeptide can then be concentrated and purified by any selected method.

Epitope-tagged PRO874 polypeptide may also be expressed in host CHO cells. The PRO874-encoding DNA may be subcloned out of the pRK5 vector. The subclone insert can undergo PCR to fuse in frame with a selected epitope tag such as a poly-his tag into a Baculovirus expression vector. The poly-his tagged PRO874-encoding DNA insert can then be subcloned into a SV40 driven vector containing a selection marker such as DHFR for selection of stable clones. Finally, the CHO cells can be transfected (as described above) with the SV40 driven vector. Labeling may be performed, as described above, to verify expression. The culture medium containing the expressed poly-His tagged PRO874 polypeptide can then be concentrated and purified by any selected method, such as by Ni^{2+}-

chelate affinity chromatography.

EXAMPLE 5: Expression of a PRO874 Polypeptide in Yeast

The following method describes recombinant expression of PRO874 polypeptides in yeast.

First, yeast expression vectors are constructed for intracellular production or secretion of PRO874 polypeptide from the ADH2/GAPDH promoter. DNA encoding the PRO874 polypeptide of interest, a selected signal peptide and the promoter is inserted into suitable restriction enzyme sites in the selected plasmid to direct intracellular expression of the PRO874 polypeptide. For secretion, DNA encoding the PRO874 polypeptide can be cloned into the selected plasmid, together with DNA encoding the ADH2/GAPDH promoter, the yeast alpha-factor secretory signal/leader sequence, and linker sequences (if needed) for expression of the PRO874 polypeptide.

Yeast cells, such as yeast strain AB110, can then be transformed with the expression plasmids described above and cultured in selected fermentation media. The transformed yeast supernatants can be analyzed by precipitation with 10% trichloroacetic acid and separation by SDS-PAGE, followed by staining of the gels with Coomassie Blue stain.

Recombinant PRO874 polypeptide can subsequently be isolated and purified by removing the yeast cells from the fermentation medium by centrifugation and then concentrating the medium using selected cartridge filters. The concentrate containing the PRO874 polypeptide may further be purified using selected column chromatography resins.

EXAMPLE 6: Expression of PRO874 Polypeptides in Baculovirus-Infected Insect Cells

The following method describes recombinant expression of PRO874 polypeptides in Baculovirus-infected insect cells.

The PRO874-encoding DNA is fused upstream of an epitope tag contained within a baculovirus expression vector. Such epitope tags include poly-his tags and immunoglobulin tags (like Fc regions of IgG). A variety of plasmids may be employed, including plasmids derived from commercially available plasmids such as pVL1393 (Novagen). Briefly, the PRO874-encoding DNA or the desired portion of the PRO874-encoding DNA (such as the sequence encoding the extracellular domain of a transmembrane protein) is amplified by PCR with primers complementary to the 5' and 3' regions. The 5' primer may incorporate flanking

(selected) restriction enzyme sites. The product is then digested with those selected restriction enzymes and subcloned into the expression vector.

Recombinant baculovirus is generated by co-transfecting the above plasmid and BaculoGold™ virus DNA (Pharmingen) into *Spodoptera frugiperda* ("Sf9") cells (ATCC CRL 1711) using lipofectin (commercially available from GIBCO-BRL). After 4 to 5 days of incubation at 28°C, the released viruses are harvested and used for further amplifications. Viral infection and protein expression is performed as described by O'Reilley et al., Baculovirus expression vectors: A laboratory Manual, Oxford: Oxford University Press (1994).

Expressed poly-his tagged PRO874 polypeptide can then be purified, for example, by Ni^{2+}-chelate affinity chromatography as follows. Extracts are prepared from recombinant virus-infected Sf9 cells as described by Rupert et al., Nature, 362:175-179 (1993). Briefly, Sf9 cells are washed, resuspended in sonication buffer (25 mL Hepes, pH 7.9; 12.5 mM $MgCl_2$; 0.1 mM EDTA; 10% Glycerol; 0.1% NP-40; 0.4 M KCl), and sonicated twice for 20 seconds on ice. The sonicates are cleared by centrifugation, and the supernatant is diluted 50-fold in loading buffer (50 mM phosphate, 300 mM NaCl, 10% Glycerol, pH 7.8) and filtered through a 0.45 μm filter. A Ni^{2+}-NTA agarose column (commercially available from Qiagen) is prepared with a bed volume of 5 mL, washed with 25 mL of water and equilibrated with 25 mL of loading buffer. The filtered cell extract is loaded onto the column at 0.5 mL per minute. The column is washed to baseline A_{280} with loading buffer, at which point fraction collection is started. Next, the column is washed with a secondary wash buffer (50 mM phosphate; 300 mM NaCl, 10% Glycerol, pH 6.0), which elutes nonspecifically bound protein. After reaching A_{280} baseline again, the column is developed with a 0 to 500 mM Imidazole gradient in the secondary wash buffer. One mL fractions are collected and analyzed by SDS-PAGE and silver staining or western blot with Ni^{2+}-NTA-conjugated to alkaline phosphatase (Qiagen). Fractions containing the eluted His_{10}-tagged PRO874 polypeptide are pooled and dialyzed against loading buffer.

Alternatively, purification of the IgG tagged (or Fc tagged) PRO874 polypeptide can be performed using known chromatography techniques, including for instance, Protein A or protein G column chromatography.

EXAMPLE 7: Preparation of Antibodies that Bind PRO874 Polypeptides

This example illustrates the preparation of monoclonal antibodies which can specifically

bind to PRO874 polypeptides.

Techniques for producing the monoclonal antibodies are known in the art and are described, for instance, in Goding, supra. Immunogens that may be employed include purified PRO874 polypeptide, fusion proteins containing a PRO874 polypeptide, and cells expressing recombinant PRO874 polypeptide on the cell surface. Selection of the immunogen can be made by the skilled artisan without undue experimentation.

Mice, such as Balb/c, are immunized with the PRO874 immunogen emulsified in complete Freund's adjuvant and injected subcutaneously or intraperitoneally in an amount from 1-100 micrograms. Alternatively, the immunogen is emulsified in MPL-TDM adjuvant (Ribi Immunochemical Research, Hamilton, MT) and injected into the animal's hind foot pads. The immunized mice are then boosted 10 to 12 days later with additional immunogen emulsified in the selected adjuvant. Thereafter, for several weeks, the mice may also be boosted with additional immunization injections. Serum samples may be periodically obtained from the mice by retro-orbital bleeding for testing in ELISA assays to detect anti-PRO874 polypeptide antibodies.

After a suitable antibody titer has been detected, the animals "positive" for antibodies can be injected with a final intravenous injection of PRO874 polypeptide. Three to four days later, the mice are sacrificed and the spleen cells are harvested. The spleen cells are then fused (using 35% polyethylene glycol) to a selected murine myeloma cell line such as P3X63AgU.1, available from ATCC, No. CRL 1597. The fusions generate hybridoma cells which can then be plated in 96 well tissue culture plates containing HAT (hypoxanthine, aminopterin, and thymidine) medium to inhibit proliferation of non-fused cells, myeloma hybrids, and spleen cell hybrids.

The hybridoma cells will be screened in an ELISA for reactivity against PRO874 polypeptide. Determination of "positive" hybridoma cells secreting the desired monoclonal antibodies against a PRO874 polypeptide is within the skill in the art.

The positive hybridoma cells can be injected intraperitoneally into syngeneic Balb/c mice to produce ascites containing the anti-PRO874 polypeptide monoclonal antibodies. Alternatively, the hybridoma cells can be grown in tissue culture flasks or roller bottles. Purification of the monoclonal antibodies produced in the ascites can be accomplished using ammonium sulfate precipitation, followed by gel exclusion chromatography. Alternatively, affinity chromatography based upon binding of antibody to protein A or protein G can be

employed.

Deposit of Material

The following materials have been deposited with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, MD, USA (ATCC):

Material	ATCC Dep. No.	Deposit Date
DNA40621-1440		June 2, 1998.

This deposit was made under the provisions of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purpose of Patent Procedure and the Regulations thereunder (Budapest Treaty). This assures maintenance of a viable culture of the deposit for 30 years from the date of deposit. The deposit will be made available by ATCC under the terms of the Budapest Treaty, and subject to an agreement between Genentech, Inc. and ATCC, which assures permanent and unrestricted availability of the progeny of the culture of the deposit to the public upon issuance of the pertinent U.S. patent or upon laying open to the public of any U.S. or foreign patent application, whichever comes first, and assures availability of the progeny to one determined by the U.S. Commissioner of Patents and Trademarks to be entitled thereto according to 35 USC §122 and the Commissioner's rules pursuant thereto (including 37 CFR §1.14 with particular reference to 886 OG 638).

The assignee of the present application has agreed that if a culture of the materials on deposit should die or be lost or destroyed when cultivated under suitable conditions, the materials will be promptly replaced on notification with another of the same. Availability of the deposited material is not to be construed as a license to practice the invention in contravention of the rights granted under the authority of any government in accordance with its patent laws.

The foregoing written specification is considered to be sufficient to enable one skilled in the art to practice the invention. The present invention is not to be limited in scope by the construct deposited, since the deposited embodiment is intended as a single illustration of certain aspects of the invention and any constructs that are functionally equivalent are within the scope of this invention. The deposit of material herein does not constitute an admission that the written description herein contained is inadequate to enable the practice of any aspect of the invention, including the best mode thereof, nor is it to be construed as limiting the scope

of the claims to the specific illustrations that it represents. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description and fall within the scope of the appended claims.

WHAT IS CLAIMED IS:

1. Isolated nucleic acid comprising DNA having at least an 80% sequence identity to (a) a DNA molecule encoding a PRO874 polypeptide comprising the sequence of amino acid residues 1 to 321 of Figure 2 (SEQ ID NO:3), or (b) the complement of the DNA molecule of (a).

2. The nucleic acid of Claim 1, wherein said DNA comprises the nucleotide sequence of SEQ ID NO:1 or its complement.

3. The nucleic acid of Claim 1, wherein said DNA comprises nucleotides 1-963 of the nucleotide sequence of SEQ ID NO:1 (SEQ ID NO:2).

4. Isolated nucleic acid comprising DNA having at least an 80% sequence identity to (a) a DNA molecule encoding the same mature polypeptide encoded by the human protein cDNA in ATCC Deposit No. ______________________ (DNA40621-1440), or (b) the complement of the DNA molecule of (a).

5. The nucleic acid of Claim 4 which comprises a DNA molecule encoding the same mature polypeptide encoded by the human protein cDNA in ATCC Deposit No. ____ ______ (DNA40621-1440).

6. Isolated nucleic acid comprising DNA having at least an 80% sequence identity to (a) a DNA molecule encoding a PRO874 polypeptide comprising the sequence of amino acid residues about X to 321 of Figure 2 (SEQ ID NO:3), or b) the complement of the DNA molecule of (a), wherein X is any one of amino acid residues 270 to 279 of Figure 2 (SEQ ID NO:3).

7. A vector comprising the nucleic acid of any one of Claims 1 to 6.

8. The vector of Claim 7 operably linked to control sequences recognized by a host cell transformed with the vector.

9. A host cell comprising the vector of Claim 7.

10. The host cell of Claim 9, wherein said cell is a CHO cell.

11. The host cell of Claim 9, wherein said cell is an *E. coli*.

12. The host cell of Claim 9, wherein said cell is a yeast cell.

13. A process for producing a PRO874 polypeptide comprising culturing the host cell of Claim 9 under conditions suitable for expression of said PRO874 polypeptide and recovering said PRO874 polypeptide from the cell culture.

14. Isolated native sequence PRO874 polypeptide comprising amino acid residues 1 to 321 of Figure 2 (SEQ ID NO:3).

15. Isolated PRO874 polypeptide comprising amino acids X to 321 of the amino acid sequence shown in Figure 2 (SEQ ID NO:3), wherein X is any one of amino acids 270 to 279 of Figure 2 (SEQ ID NO:3).

16. Isolated PRO874 polypeptide encoded by the cDNA insert of the vector deposited as ATCC Accession No. ______________ (DNA40621-1440).

17. A chimeric molecule comprising a PRO874 polypeptide fused to a heterologous amino acid sequence.

18. The chimeric molecule of Claim 17, wherein said heterologous amino acid sequence is an epitope tag sequence.

19. The chimeric molecule of Claim 17, wherein said heterologous amino acid sequence is a Fc region of an immunoglobulin.

20. An antibody which specifically binds to a PRO874 polypeptide.

21. The antibody of Claim 20, wherein said antibody is a monoclonal antibody.

Abstract of the Disclosure

The present invention is directed to novel multi-span transmembrane polypeptides and to nucleic acid molecules encoding those polypeptides. Also provided herein are vectors and host cells comprising those nucleic acid sequences, chimeric polypeptide molecules comprising the polypeptides of the present invention fused to heterologous polypeptide sequences, antibodies which bind to the polypeptides of the present invention and to methods for producing the polypeptides of the present invention.

FIGURE 1A

```
  1 CGGACGCGTG GGCGGACGCG TGGGGGCTGT GAGAAAGTGC CAATAAATAC ATCATGCAAC CCCACGGCCC ACCTTGTGAA CTCCTCGTGC CCAGGGCTGA
    GCCTGCGCAC CCGCCTGCGC ACCCCCGACA CTCTTTCACG GTTATTTATG TAGTACGTTG GGGTGCCGGG TGGAACACTT GAGGAGCACG GGTCCCGACT
  1 R  T  R  G   R  T  R   G  G  C   E  K  V  P   I  N  T   S  C  N   P  T  A  H   L  V  N   S  S  C   P  G  L  M
    ^ORF                                                                                                            ^Met

101 TGTGCGTCTT CCAGGGCTAC TCATCCAAAG GCCTAATCCA ACGTTCTGTC TTCAATCTGC AAATCTATGG GGTCCTGGGG CTCTTCTGGA CCCTTAACTG
    ACACGCAGAA GGTCCCGATG AGTAGGTTTC CGGATTAGGT TGCAAGACAG AAGTTAGACG TTTAGATACC CCAGGACCCC GAGAAGACCT GGGAATTGAC
 35   C  V  F   Q  G  Y   S  S  K  G   L  I  Q   R  S  V   F  N  L  Q   I  Y  G   V  L  G   L  F  W  T   L  N  W

201 GGTACTGGCC CTGGGCCAAT GCGTCCTCGC TGGAGCCTTT GCCTCCTTCT ACTGGGCCTT CCACAAGCCC CAGGACATCC CTACCTTCCC CTTAATCTCT
    CCATGACCGG GACCCGGTTA CGCAGGAGCG ACCTCGGAAA CGGAGGAAGA TGACCCGGAA GGTGTTCGGG GTCCTGTAGG GATGGAAGGG GAATTAGAGA
 68 V  L  A   L  G  Q  C   V  L  A   G  A  F   A  S  F  Y   W  A  F   H  K  P   Q  D  I  P   T  F  P   L  I  S

301 GCCTTCATCC GCACACTCCG TTACCACACT GGGTCATTGG CATTTGGAGC CCTCATCCTG ACCCTTGTGC AGATAGCCCG GGTCATCTTG GAGTATATTG
    CGGAAGTAGG CGTGTGAGGC AATGGTGTGA CCCAGTAACC GTAAACCTCG GGAGTAGGAC TGGGAACACG TCTATCGGGC CCAGTAGAAC CTCATATAAC
101 A  F  I  R   T  L  R   Y  H  T   G  S  L  A   F  G  A   L  I  L   T  L  V  Q   I  A  R   V  I  L   E  Y  I  D

401 ACCACAAGCT CAGAGGAGTG CAGAACCCTG TAGCCCGCTG CATCATGTGC TGTTTCAAGT GCTGCCTCTG GTGTCTGGAA AAATTTATCA AGTTCCTAAA
    TGGTGTTCGA GTCTCCTCAC GTCTTGGGAC ATCGGGCGAC GTAGTACACG ACAAAGTTCA CGACGGAGAC CACAGACCTT TTTAAATAGT TCAAGGATTT
135   H  K  L   R  G  V   Q  N  P  V   A  R  C   I  M  C   C  F  K  C   C  L  W   C  L  E   K  F  I  K   F  L  N

501 CCGCAATGCA TACATCATGA TCGCCATCTA CGGGAAGAAT TTCTGTGTCT CAGCCAAAAA TGCGTTCATG CTACTCATGC GAAACATTGT CAGGGTGGTC
    GGCGTTACGT ATGTAGTACT AGCGGTAGAT GCCCTTCTTA AAGACACAGA GTCGGTTTTT ACGCAAGTAC GATGAGTACG CTTTGTAACA GTCCCACCAG
168  R  N  A   Y  I  M  I   A  I  Y   G  K  N   F  C  V  S   A  K  N   A  F  M   L  L  M  R   N  I  V   R  V  V

601 GTCCTGGACA AAGTCACAGA CCTGCTGCTG TTCTTTGGGA AGCTGCTGGT GGTCGGAGGC GTGGGGGTCC TGTCCTTCTT TTTTTTCTCC GGTCGCATCC
    CAGGACCTGT TTCAGTGTCT GGACGACGAC AAGAAACCCT TCGACGACCA CCAGCCTCCG CACCCCCAGG ACAGGAAGAA AAAAAAGAGG CCAGCGTAGG
201 V  L  D  K   V  T  D   L  L  L   F  F  G  K   L  L  V   V  G  G   V  G  V  L   S  F  F   F  F  S   G  R  I  P

701 CGGGGCTGGG TAAAGACTTT AAGAGCCCCC ACCTCAACTA TTACTGGCTG CCCATCATGA CCTCCATCCT GGGGGCCTAT GTCATCGCCA GCGGCTTCTT
    GCCCCGACCC ATTTCTGAAA TTCTCGGGGG TGGAGTTGAT AATGACCGAC GGGTAGTACT GGAGGTAGGA CCCCCGGATA CAGTAGCGGT CGCCGAAGAA
235   G  L  G   K  D  F   K  S  P  H   L  N  Y   Y  W  L   P  I  M  T   S  I  L   G  A  Y   V  I  A  S   G  F  F

801 CAGCGTTTTC GGCATGTGTG TGGACACGCT CTTCCTCTGC TTCCTGGAAG ACCTGGAGCG GAACAACGGC TCCCTGGACC GGCCCTACTA CATGTCCAAG
    GTCGCAAAAG CCGTACACAC ACCTGTGCGA GAAGGAGACG AAGGACCTTC TGGACCTCGC CTTGTTGCCG AGGGACCTGG CCGGGATGAT GTACAGGTTC
268  S  V  F   G  M  C  V   D  T  L   F  L  C   F  L  E  D   L  E  R   N  N  G   S  L  D  R   P  Y  Y   M  S  K
```

FIGURE 1B

```
 901 AGCCTTCTAA AGATTCTGGG CAAGAAGAAC GAGGCGCCCC CGGACAACAA GAAGAGGAAG AAGTGACAGC TCCGGCCCTG ATCCAGGACT GCACCCCACC
     TCGGAAGATT TCTAAGACCC GTTCTTCTTG CTCCGCGGGG GCCTGTTGTT CTTCTCCTTC TTCACTGTCG AGGCCGGGAC TAGGTCCTGA CGTGGGGTGG
 301 S  L  L  K   I  L  G   K  K  N   E  A  P  P   D  N  K   K  R  K   K  O
                                                                            ^40621.tm.f1

1001 CCCACCGTCC AGCCATCCAA CCTCACTTCG CCTTACAGGT CTCCATTTTG TGGTAAAAAA AGGTTTTAGG CCAGGCGCCG TGGCTCACGC CTGTAATCCA
     GGGTGGCAGG TCGGTAGGTT GGAGTGAAGC GGAATGTCCA GAGGTAAAAC ACCATTTTTT TCCAAAATCC GGTCCGCGGC ACCGAGTGCG GACATTAGGT
     ^40621.tm.p1                                ^40621.tm.r1

1101 ACACTTTGAG AGGCTGAGGC GGGCGGATCA CCTGAGTCAG GAGTTCGAGA CCAGCCTGGC CAACATGGTG AAACCTCCGT CTCTATTAAA AATACAAAAA
     TGTGAAACTC TCCGACTCCG CCCGCCTAGT GGACTCAGTC CTCAAGCTCT GGTCGGACCG GTTGTACCAC TTTGGAGGCA GAGATAATTT TTATGTTTTT

1201 TTAGCCGAGA GTGGTGGCAT GCACCTGTCA TCCCAGCTAC TCGGGAGGCT GAGGCAGGAG AATCGCTTGA ACCCGGGAGG CAGAGGTTGC AGTGAGCCGA
     AATCGGCTCT CACCACCGTA CGTGGACAGT AGGGTCGATG AGCCCTCCGA CTCCGTCCTC TTAGCGAACT TGGGCCCTCC GTCTCCAACG TCACTCGGCT

1301 GATCGCGCCA CTGCACTCCA ACCTGGGTGA CAGACTCTGT CTCCAAAACA AAACAAACAA ACAAAAAGAT TTTATTAAAG ATATTTTGTT AACTC
     CTAGCGCGGT GACGTGAGGT TGGACCCACT GTCTGAGACA GAGGTTTTGT TTTGTTTGTT TGTTTTTCTA AAATAATTTC TATAAAACAA TTGAG
```

FIGURE 2

```
1       10        20        30        40        50        60        70
|    |    |    |    |    |    |    |    |    |    |    |    |    |
RTRGRTRGGCEKVPINTSCNPTAHLVNSSCPGLMCVFQGYSSKGLIQRSVFNLQIYGVLGLFWTLNWVLA

71      80        90        100       110       120       130       140
|    |    |    |    |    |    |    |    |    |    |    |    |    |
LGQCVLAGAFASFYWAFHKPQDIPTFPLISAFIRTLRYHTGSLAFGALILTLVQIARVILEYIDHKLRGV

141     150       160       170       180       190       200       210
|    |    |    |    |    |    |    |    |    |    |    |    |    |
QNPVARCIMCCFKCCLWCLEKFIKFLNRNAYIMIAIYGKNFCVSAKNAFMLLMRNIVRVVVLDKVTDLLL

211     220       230       240       250       260       270       280
|    |    |    |    |    |    |    |    |    |    |    |    |    |
FFGKLLVVGGVGVLSFFFFSGRIPGLGKDFKSPHLNYYWLPIMTSILGAYVIASGFFSVFGMCVDTLFLC

281     290       300       310       320
|    |    |    |    |    |    |    |    |
FLEDLERNNGSLDRPYYMSKSLLKILGKKNEAPPDNKKRKK
```

FEATURES:

Transmembrane domains

57-80	GVLGLFWTLNWVLALGQCVLAGAF	(Type II)
110-126	TGSLAFGALILTLVQIA	
215-231	LLVVGGVGVLSFFFFSG	
254-274	TSILGAYVIASGFFSVFGMCV	

Potential N-glycosylation sites

16-19	NTSC
27-30	NSSC
289-292	NGSL

Sequence Identity with Hypothetical YBR002c family proteins: 276-287 TLFLCFLEDLER

Sequence Identity with Ammonium transporter proteins: 204-230 KVTDLLLFFGKLLVVGGVGVLSFFFFS

FIGURE 3A

```
714306      1 GGGCCCCTGGTGCTGGTGTGATCCTGGGAGTGCTGGGCGTGCTGGCATAT
2743044     1                 GCTGATCCTGGGAGTGCTGGGCGTGCTGGCATAT
                +++++++++++++++..+++++++++++++++++++++++++++++++++
consen01    1 GGGCCCCTGGTGCTGGTGTGATCCTGGGAGTGCTGGGCGTGCTGGCATAT

714306     51 GGCATCTACTACTGCTGGGAGGAGTACCGAGTG-TGCGGGACAAGGGCGC
2743044    35 GGCATCTACTACTGCTGGGAGGAGTACCGAGTGCTGCGGGACAAGGGCGC
              ++++++++++++++++++++++++++++++++++++++++++++++++++
consen01   51 GGCATCTACTACTGCTGGGAGGAGTACCGAGTGCTGCGGGACAAGGGCGC

714306    100 CTCCATCTCCCAGCTGGGTTTCACCACCAAC-TCAGTGCCTACCAGAGCG
2743044    85 CTCCATCTCCCAGCTGGGTTTCACCACCAACCTCAGTGCCTACCAGAGCG
1736230     1                                 CTCAGTGCCTACCAGAGCG
              ++++++++++++++++++++++++++++++++++++++++++++++++++
consen01  101 CTCCATCTCCCAGCTGGGTTTCACCACCAACCTCAGTGCCTACCAGAGCG

714306    149 TGCAGGAGACCTGGCTGGCCGCCCTGATCGTGTTGGCGGTGCTTGAAGCC
2743044   135 TGCAGGAGACCTGGCTGGCCGCCCTGATCGTGTTGGCGGTGCTTGAAGCC
1736230    20 TGCAGGAGACCTGGCTGGCCGCCCTGATCGTGTTGGCGGTGCTTGAAGCC
1807819     1            TGGCTGGCCGCCCTGATCGTGTTGGCGGTGCTTGAAGCC
841737      1                                              AAGCC
              ++++++++++++++++++++++++++++++++++++++++++++++++++
consen01  151 TGCAGGAGACCTGGCTGGCCGCCCTGATCGTGTTGGCGGTGCTTGAAGCC

714306    199 ATCCTGCTGCTGGTGCTCATCTTCCTGCGGCAGCGGA-TTCGTATTGCCA
2743044   185 ATCCTGCTGCTGGTGCTCATCTTCCTGCGGCAN-GGA-TTCGTATTGCCA
1736230    70 ATCCTGCTGCTGATGCTCATCTTCCTGCGGCA-CGGANTTCGTATTGC
1807819    40 ATCCTGCTGCTGATGCTCATCTTCCTGCGGCAG-GGA-TTCGTATTGCCA
841737      6 ATNCTGCTGCTGATGNTNATNTTNCTGCGGCAG-GGA-TTCGTATTGCCA
2627821     1 ATCCTGCTGCTGATGCTCATCTTCCTGCGGCAGCGGA-TTCGTATTGCCA
2679739     1                                             TTGCCA
              ++.+++++++++ ++.+.++.++.++++++++.++++ ++++++++++++
consen01  201 ATCCTGCTGCTGGTGCTCATCTTCCTGCGGCAGCGGA TTCGTATTGCCA

714306    248 TCGCCCTCCTGAAGGAGGCCAGCAAGGCTGTGGGACAGATGAT
2743044   233 TCGCCCTCC
1807819    88 TCGCCCTCCTGAAGGAGGCCAGCAAGGCTGTGGGACAGATGATGTCTACC
841737     54 TCGGCCTNCTGAAGGAGGCCAGCAAGGCTGTGGGACAGATGATGTCTACC
2627821    50 TCGCCCTCCTGAAGGAGGCCAGCAAGGCTGTGGGACAGATGATGTCTACC
2679739     7 TCGCCCTCCTGAACGANGCCAGCAAGGCTGTGGGACAGATGATGTCTACC
2680211     1          CTGAAGGAGGCCAGCAAGGCTGTGGGACAGATGATGTCTACC
              +++.+++.+++++.++.+++++++++++++++++++++++++++++++++
consen01  250 TCGCCCTCCTGAAGGAGGCCAGCAAGGCTGTGGGACAGATGATGTCTACC

1807819   138 ATGT
841737    104 ATGTTCTACCCACTGGTCAACTTTGTCCTNCTTCTCATCTGCATTGCCTA
2627821   100 ATGTTCTACCCACTGGTCACCTTTGTCCTCCTCCTCATCTGCATTGCCTA
2679739    57 ATGTTCTACCCACTGGTCACCTTTGTCCTCCTCCTCATCTGCATTGCCTA
2680211    43 ATGTTCTACCCACTGGTCACCTTTGTCCTCCTCCTCATCTGCATTGCCTA
2529715     1                   CCTTTGTCCTCCTCCTCATCTGCATTGCCTA
              ++++++++++++++++++.+++++++++.++.++++++++++++++++
consen01  300 ATGTTCTACCCACTGGTCACCTTTGTCCTCCTCCTCATCTGCATTGCCTA

841737    154 -CTGGGCCATGACTGCTCTGTANCTGGGTACATCGGGGCAACCCCAGTAT
2627821   150 -CTGGGCCATGACTGCTCT-TACCTGGCTACATCGGGGCAACCCCAGTAT
2679739   107 -CTGGGCCATGACTGCTCTGTACCTGGCTACATCGGGGCAACCCCAGTAT
```

FIGURE 3B

```
2680211             93 -CTGGGCCATGACTGCTCTGTACCTGGCTACATCGGGGCAACCCCAGTAT
2529715             32 ACTGGGCCATGACTGCTCTGTACCTGGCTACATCGGGGCAACCCCANTAT
835107               1  CTGGGNCACGAGTGNTCTGTACCTGGNTACATCGGGGCAACCCCAGTAT
1692719              1                          CGGCTCGAGCGGGGCAACCCCAGTAT
                        +++++.++.++.++.++++++.+.++.+..+.+++++++++++++.+++
consen01           350  CTGGGCCATGACTGCTCTGTACCTGGCTACATCGGGGCAACCCCAGTAT

841737             203 GTGCTCTGGGNATNCAACATC-AGNT-NCCCCGGCTGTGAGAAAGTGC
2627821            198 GTGCTCTGGGCATCCAACATC-AGCT-CCCCCGGCTGTGAGA
2679739            156 GTGCTCTGGGCATCCAACATC-AGCT-CCCCCGGCTGTGAGAAAGTGCCA
2680211            142 GTGCTCTGGGCATCCAACATC-AGCT-CCCCCGGCTGTGAGAAAGTGCCA
2529715             82 GTGCTCTGGGCATCCAACATCCAGCTNCCCCCGGCTGTGAGAAAGTGCCA
835107              50 GTGCTCTGGGCATCCAACATC-AGCT-CCCCCGGCTGTGAGAAAGTGCCA
1692719             27 GTGCTCTGGGCATCCAACATC-AGCT-CCCCCGGCTGTGAGAAAGTGCCA
                       ++++++++++.++.+++++++ ++.+ .++++++++++++++++++++++
consen01           399 GTGCTCTGGGCATCCAACATC AGCT CCCCCGGCTGTGAGAAAGTGCCA

2679739            204 ATAAATACATCATGCAACCCCACGGCCCACCTTGTGAACTCCTCGTGCCC
2680211            190 ATAAATACATCNTGCAACCCCACGGCCCACCTTGTGAACTCCTCGTGCCC
2529715            132 ATAAATACATCATGCAACCCCACGGCCCACCTTGTGAACTCCTCGTGCCC
835107              98 -TAAATACATCATGCAACCCCACGGCCCACCTTGTGAACTCCTCGTGCCC
1692719             75 ATAAATACATCATGCAACCCCACGGCCCACCTTGTGAACTCCTCGTGCCC
3210594              1       ACATCATGCAACCCCACGGCCCACCTTGTGAACTCCTCGTGCCC
2264423              1             TGCAACCCCACGGCCCACCTTGTGAACTCCTCGTGCCC
                       +++++++++++.++++++++++++++++++++++++++++++++++++++
consen01           447 ATAAATACATCATGCAACCCCACGGCCCACCTTGTGAACTCCTCGTGCCC

2679739            254 AGGG-CTGATGTGCGTCTTCCA
2680211            240 ANGG-CTGATGTGCGTCTTCCAGGGCTACTCATCCAAAGGCCTAATCCAA
2529715            182 AGGG-CTGATGTGCGTCTTCCAGGG-TACTCATCCAAAGGCCTAATCCAA
835107             147 AGGG-CTGATGTGCGTCTTCCAGGGCTACTCATCCAAAGGCCTAATCCAA
1692719            125 AGGG-CTGATGTGCGTCTTCCAGGGCTACTCATCCAAAGGCCTAATCCAA
3210594             45 AGGGGCTGATGTGCGTCTTCCAGGGCTACTCATCCAAAGGCCTAATCCAA
2264423             39 AGGG-CTGATGTGCGTCTTCCAGGGCTACTCATCCAAAGGCCTAATCCAA
1710447              1                             CTCATCCAAAGGCCTAATCCAA
1727529              1                                                CAA
                       +.++++++++++++++++++++++++++++++++++++++++++++++++
consen01           497 AGGGGCTGATGTGCGTCTTCCAGGGCTACTCATCCAAAGGCCTAATCCAA

2680211            289 CGTTCTGTCT
2529715            230 CGTTCTGTCTTCAATCNGCAAATCTATGGGGTCCTGGGGCA
835107             196 CGTTCTGTCTTCAATCTGCAAATCTATGGGGTCCTGGGGCTCTTCTGGAC
1692719            174 CGTTCTGTCTTCAATCTGCAAATCTATGGGGTCCTGGGGCTCTTCTG
3210594             95 CGTTCTGTCTTCAATCTGCAAATCTATGGNGTCCTGGGGCTCTTCTGGAC
2264423             88 CGTTCTGTCTTCAATCTGCAAATCTATGGNGTCCTGGGGCTCTTCTGGAC
1710447             23 CGTTCTGTCTTCAATCTGCAAATCTATGGNGTCCTGGGGCTCTTCTGGAC
1727529              4 CGTTCTGTCTTCAATCTGCAAATCTATGGNGTCCTGGGGCTCTTCTGGAC
1858570              1                          ATGGGGTCCTGGGGCTCTTCTGGAC
1797994              1                                   TNGGGCTCTTCTGGAC
1689486              1                                        CTCTTCTGGAC
DNA36459.init        1                                        CTCTTCTGGAC
                       ++++++++++++++++.+++++++++++.+++++.++++.++++++++++
consen01           547 CGTTCTGTCTTCAATCTGCAAATCTATGGGGTCCTGGGGCTCTTCTGGAC

835107             246 CCTTAACTGGGTACTGGCCCTGGGCCAATG
3210594            145 CCTTAACTGGGTACTGGCCCTGGGCCAATGCGTCCTCGCTGGAGCCTTTG
2264423            138 CCTTAACTGGGTACTGGCCCTGGGCCAATGCGTCCTCGCTGGAGCCTTTG
```

FIGURE 3C

```
1710447          73  CCTTAACTGGGTACTGGCCCTGGGCCAATGCGTCCTCGCTGGAGCCTTTG
1727529          54  CCTTAACTGGGTACTGGCCCTGGGCCAATGCGTCCTCGCTGGAGCCTTTG
1858570          26  CCTTAACTGGGTACTGGCCCTGGGCCAATGCGTCCTCGCTGCAGCCTTTG
1797994          17  CCTTAACTGGGTACTGGCCCTGGGCCAATGCGTCCTCGCTGGAGCCTTTG
1689486          12  CCTTAACTGGGTACTGGCCCTGGGCCAATGCGTCCTCGCTGGAGCCTTTG
DNA36459.init    12  CCTTAACTGGGTACTGGCCCTGGGCCAATGCGTCCTCGCTGGAGCCTTTG
898598            1                                                   G
                     ++++++++++++++++++++++++++++++++++++++++.+++++++++
consen01        597  CCTTAACTGGGTACTGGCCCTGGGCCAATGCGTCCTCGCTGGAGCCTTTG

3210594         195  CCTCCTTCTACTGGCCTTCCACAAGCCCCAGGACATCCCTACCTTCCCC
2264423         188  CCTCCTTCTACTGGCCTTCCACAAGCCCCAGGACATCCCTACCTTCCCC
1710447         123  CCTCCTTCTACTGGCCTTCCACAAGCCCCAGGACATCCCTACCTTCCCC
1727529         104  CCTCCTTCTACTGGCCTTCCACAAGCCCCAGGACATCCCTACCTTCCCC
1858570          76  CCTCCTTCTACTGGCCTTCCACAAGCCCCAGGACATCCCTACCTTCCCC
1797994          67  CCTCCTNCTACTGGCCTTCCACAAGCCCCAGGACATCCCTNCTTNCCCN
1689486          62  CCTCCTTCTACTGGCCTTCCACAAGCCCCAGGACATCCCTACCTTCCCC
DNA36459.init    62  CCTCCTTCTACTGGCCTTCCACAAGCCCCAGGACATCCCTACCTTCCCC
898598            2  CCTCCTTCTACTGGCCTTCCACAAGCCCCAGGACATCCCTACCTTCCCC
                     ++++++.+++++++++++++++++++++++++++++++++.+.+.+++.
consen01        647  CCTCCTTCTACTGGCCTTCCACAAGCCCCAGGACATCCCTACCTTCCCC

3210594         245  TTAATCTCTGCCTTCATCCGCACACTCCGTNACCACACTGGGTCA
2264423         238  TTAATCTCTGCCT
1710447         173  TTAATCTCTGCCTTCATCCGCACACTCCGTTACCACACTAA
1727529         154  TTAATCTCTGCCTTCATCCGCACACTCCGTTACCACACTGGGTCAT
1858570         126  TTAATCTCTGCCTTCATCCGCACACTCCGTTACCACACTGGGTCATTGGC
1797994         117  TAANTCTCTGCNTTCATCCGCACACTCCGTTACCACACTGGGTCATTGGC
1689486         112  TTAATCTCTGCCTTCATCCGCACACTCCGTTACCACACTGGGTCATTGGC
DNA36459.init   112  TTAATCTCTGCCTTCATCCGCACACTCCGTTACCACACTGGGTCATTGGC
898598           52  TTAATCTCTGCCTTCATCCGCACACTCCGTTACCACACTGGGTCATTGGC
785379            1                       CACACTCCGTTACCACACTGGGTCATTGGC
1859705           1                                                ATTGGC
                     +.+.+++++++.+++++++++++++++++.++++++++..+++++++++
consen01        697  TTAATCTCTGCCTTCATCCGCACACTCCGTTACCACACTGGGTCATTGGC

1858570         176  ATTTGGAGCCCTCATCCTGACCCTTGTGCAGATAGCCCGGGTCATCTTGG
1797994         167  ATTTGGAGCCNTCATCCTGGACCTTGTGNNGATAGCCCGGGGTCG
1689486         162  ATTTGGAGCCCTCATCCTGACCCTTGTGCAGATAGCCCGGGTCATCTTGG
DNA36459.init   162  ATTTGGAGCCCTCATCCTGACCCTTGTGCAGATAGCCCGGGTCATCTTGG
898598          102  ATTTGGAGCCCTCATCCTGACCCTTGTGCAGATAGCCCGGGTCATCTTGG
785379           31  ATTTGGAGCCCTCATCCTGACCCTTGTGCAGATAGCCCGGGTCATCTTGG
1859705           7  ATTTGGAGCNCTCATCCTGACCCTTGTGCAGATAGCCCGGGTCATCTTGG
2353491           1      GGAGCCCTCATCCTGACCCTTGTGCAGATAGCCCGGGTCATCTTGG
1347151           1                 CCTGACCCTTGTGCAGATAGCCCGGGTCATCTTGG
839184            1                              GATAGCCCGGGTCATCTTGG
                     +++++++++..++++++++..+++++++..+++++++++++....+++++
consen01        747  ATTTGGAGCCCTCATCCTGACCCTTGTGCAGATAGCCCGGGTCATCTTGG

1858570         226  AGTANATTGACCACAAG-TCAGAGGAGTGCAGAACCCTGTAGCCCGCTGC
1689486         212  AGTAT
DNA36459.init   212  AGTAT
898598          152  AGTATATTGACCACAAGCTCAGAGGAGTGCAGAACCCTGTAGCCCGCTGC
785379           81  AGTATATTGACCACAAGCTCAGAGGAGTGCAGAACCCTGTAGCCCGCTGC
1859705          57  AGTATATTGACCACAAGCTCAGAGGAGTGCAGAACCCTGTAGCCCGCTGC
2353491          47  AGTATATTGACCACAAGCTCAGAGGAGTGCAGAACCCTGTAGCCCGCTGC
1347151          36  AGNATATTGACCACAAGCTCAGANGAGTGCAGANCCCTGTAGCCCGCTGC
```

FIGURE 3D

```
839184       21 NGTATATTGNCCACAAGCTCAGAGGAGTGCAGAACCCTGTAGCCCGCTGC
1548970       1           CCACAAGCTCAGAGGAGTGCAGAACCCTGTAGCCCGCTGC
                .+.+.++++.++++++++++++.+++++++++.+++++++++++++++++
consen01    797 AGTATATTGACCACAAGCTCAGAGGAGTGCAGAACCCTGTAGCCCGCTGC

1858570     275 ATCATGTG-TGTTTCAAGT
898598      202 ATCATGTGCTGTTTCAAGTGCTGCCTCTGGTGTCTGGAAAAATTTATCAA
785379      131 ATCATGTGCTGTTTCAAGTGCTGCCTCTGGTGTCTGGAAAAATTTATCAA
1859705     107 ATCATGTGCTGTTTCAAGTGCTGCCTCTGGTGTCTGGAAAAATTTATCAA
2353491      97 ATCATGTGCTGTTTCAAGTGCTGCCTCTGGTGTCTGGAAAAATTTATCAA
1347151      86 ATCATGTGCTGTTTCAAGTGCTGCCTCTGGTGTCTGGNAAANNTTATCAA
839184       71 ATCATGTGCTGTTTCAAGTGCTGCCTNTGGTGTCTGGANAAATTTATCAA
1548970      41 ATCATGTGCTGTTTCAAGTGCTGCCTCTGGTGTCTGGAAAAATTTATCAA
875537        1                  GTGCTGCCTCTGGTGTCTGGAAAAATTTATCAA
3257651       1                                     GAAAAATTTATCAA
                ++++++++++++++++++++++++++.++++++++++..++..+++++++
consen01    847 ATCATGTGCTGTTTCAAGTGCTGCCTCTGGTGTCTGGAAAAATTTATCAA

898598      252 GTTCCTAAACCGCAATGCATACATCATGATCGCCATCTACGGG
785379      181 GTTCCTAAACCGNAATGCATACATCATGATCGCCATCTACGGGAAGAATT
1859705     157 GTTCCTAAACCGCAATGCATACATCATGATCGCCATCTACGGGAAGAATT
2353491     147 GTTCCTAAACCGCAATGCATACATCATGATCGCCATCTACGGGAAGAATT
1347151     136 GTTCCTAAACCGCAATGCATACATCATGATCGCCATCTACGGGAAGANTT
839184      121 GTTCCTAAANCGCAATGCATACATCATGNTCGCCATCTACGGGAAGANTT
1548970      91 GTTCCTAAACCGCAATGCATACATCATGATCGCCATCTACGGGAAGAATT
875537       34 GTTCCTAAACCGCAATGCATACATCATGATCGCCATCTACGGGAAGAATT
3257651      15 GTTCCTAAACCGCAATGCATACATCATGATCGCCATCTACGGGAAGAATT
R24141        1                             ATCGCCATCTACGGGAAGAATT
2699329       1                              TCGCCATCTACGGGAAGAATT
3184889       1                                              GAATT
816733        1                                                ATT
                +++++++++.++.+++++++++++++++.++++++++++++++++++.++
consen01    897 GTTCCTAAACCGCAATGCATACATCATGATCGCCATCTACGGGAAGAATT

785379      231 TCTGTGTCTCAGCCAAAAATGCGTTCATGCTACTCATGCGAAACATTGTC
1859705     207 TCTGTGTCTCAGCCAAAAATGCGTTCATGCTACTCATGCGAAACATTGTC
2353491     197 TCTGTGTCTCAGCC
1347151     186 TCTGTGTCTCAGCCAAAAATGCGTTCATGNTACTCATGCGAAACA
839184      171 TCTGTGTCTCAGCCAAAAATGCGTTCATGNTACTCATGCGAAACATTGTC
1548970     141 TCTGTGTCTCAGCCAAAAATGCGTTCATGCTACTCATGCGAAACATTG
875537       84 TCTGTGTCTCAGCCAAAAATGCGTTCATGCTACTCATGCGAAACATTGTC
3257651      65 TCTGTGTCTCAGCCAAAAATGCGTTCATGCTACTCATGCGAAACATTGTC
R24141       23 TCTGTGTCTCAGCCAAAAATGCGTTCATGCTACTCATGCGAAACATTGTC
2699329      22 TCTGTGTCTCAGCCAAAAATGCGTTCATGCTACTCATGCGAAACATTGTC
3184889       6 TCTGTGTCTCAGCCAAAAATGCGTTCATGCTACTCATGCGAAACATTGTC
816733        4 TCTGTGTCTCAGCCAAAAATGCGTTCATGCTACTGATGCGAAACATTGTC
1738719       1                                    ATGCGAAACATTGTC
                +++++++++++++++++++++++++++++.++++.+++++++++++++++
consen01    947 TCTGTGTCTCAGCCAAAAATGCGTTCATGCTACTCATGCGAAACATTGTC

785379      281 AGGGTGGTCGTCCTGGAC
1859705     257 AGGGTGGTTNACC
839184      221 AGGGTGGTCGTC-TGGACAAAGTCACAGANCTGNTGCTGTTCTTTGG
875537      134 AGGGTGGTCGTCCTGGACAAAGTCACAGACCTGCTGCTGTTCTTTGGGAA
3257651     115 AGGGTGGTCGTCCTGGACAAAGTCACAGACCTGNTGCNGNTCTTTGGGAA
R24141       73 AGGGTGGTCGTCCTGGACAAAGTCACAGACCTGCTGCTGTTCTTTGGGAA
2699329      72 AGGGTGGTCGTCCTGGACAAAGTCACAGACCTGCTGCTGTTCTTTGGGAA
```

FIGURE 3E

```
3184889       56  AGGGTGGTCGTCCTGGACAAAGTCACAGACCTGCTGCTGTTCTTTGGGAA
816733        54  AGGGTGGTCGTACTGGACAAAGTCACAGANCTGCTGCTGTTCTTTGGGAA
1738719       16  AGGGTGGTCGTCCTGGACAAAGTCACAGACCTGCTGCTGTTCTTTGGGAA
1798567        1   GGGTGGTCGTCCTGGACAAAGTCACAGACCTGCTGCTGTTCTTTGGGAA
2700889        1          GTCCTGGACAAAGTCACAGACCTGCTGCTGTTCTTTGGGAA
616205         1           CCTGGACAAAGTCACAGACCTGCTGCTGTTCTTTGGGAA
615116         1             CTGGACAAAGTCACAGACCTNCTGCTGTTCTTTGGGAA
1843426        1                           ACCTGCTGCTGTTCTTTGGGAA
                  ++++++++......++++++++++++++.++..+++.+.++++++++++
consen01     997  AGGGTGGTCGTCCTGGACAAAGTCACAGACCTGCTGCTGTTCTTTGGGAA

875537       184  GCTGCTGGTGGTCGGAGGCGTGGGGGTCCTGTCCTTCTTTTTTT-CTCC
3257651      165  GCTGCTGGTGGTCNGAGGCGTGGGGGTCCTGTCNTTCTNTNTTTTCTCC
R24141       123  GCTGCTGGTGGTCGGAGGCGTGGGGGTCCTGTCCTTCTTTTTTTTCTCC
2699329      122  GCTGCTGGTGGTCGGAGGCGTGGGGGTCCTGTCCTTCTTTTTTT-CTCC
3184889      106  GCTGCTGGTGGTCGGAGGCGTGGGGGTCCTGTCCTTCTTTTTTT-CTCC
816733       104  GNTGCTGGTGGTCGGAGGCGTGGGGNACTNNACTNCTNTTTTT-CTAC
1738719       66  GCTGCTGGTGGTCGGANGCGTGGGGGTCCTGTCCTTCTTTTTTT-CTCC
1798567       50  GCTGCTGGTGGTCGGAGGCGTGGGGGTCCTGTCCTTCTTTTTTT-CTCC
2700889       42  GCTGCTGGTGGTCGGAGGCGTGGGGGTCCTGTCCTTCTTTTTTTTCTCC
616205        41  GCTGCTGGTGGTCGGAGGCGTGGGGGTCCTGTCCTTNTTTTTTT-CTCC
615116        39  GNTNCTGGTGGTNGGAGGCGTGGGGGTNCTGTCCTTCTTTTTTT-CTNC
1843426       23  GCTGCTGGTGGTCGGAGGCGTGGGGGTCCTGTCCTTCTTTTTT--CTCC
794400         1                     GGCGTGGGGGTCCTGTCCTTNTTTTTTT-CTTC
2850178        1                       CGTGGGGGTCCTGTCCTTCTTTTTTT-CTCC
1795703        1                                GTCCTTCTTTTTTT-CTCC
                  +.+.++++++++..++.+++++++++..++....+..+.+.+++++++.+
consen01    1047  GCTGCTGGTGGTCGGAGGCGTGGGGGTCCTGTCCTTCTTTTTTTTCTCC

875537       233  GGTCGCATCCCGGGGCTGGGTAAAGACTTTAAGAGCCCC
3257651      215  GGTCGCATCCCGGGGCTGGGTAAAGACTTTAAAA
R24141       173  GGTCGCATCCCGGGGCTGGGTAAAGACTTTAAGAGCCCCCACCTCAACTA
2699329      171  GGTCGCATCCCGGGGCTGGGTAAAGACTTTAAGAGCCCCCACCTCAACTA
3184889      155  GGTCACATCCCGGGGCTGGGTAAAGACTTTAAGAGCCCCCACCTCAACTA
816733       153  NNTCG-ATTCCNNGGCTGGGTAAAGACTTTAAGAGCCCCCANCTTAACTA
1738719      115  GGTCGCATCCCGGGGCTGGGTAAAGACTTTAAGAGCCCCCACCTCAACTA
1798567       99  GGTCGCATCCCGGGGCTGGGTAAAGACTTTAAGAGCCCCCACCTCAACTA
2700889       92  GGTCGCATCCCGGGGCTGGGTAAAGACTTTAAGAGCCCCCACCTCAACTA
616205        90  GGTCGGATCCCGGGGCTGGGTAAAGACTTTAAGAGCCCCCACCTCAACTA
615116        88  GGTCGGATNCCGGGGCTGGGTAAAGACTTTAAGAGNCCCCANCTNAACTA
1843426       71  GGTCGCATCCCGGGGCTGGGTAAAGACTTTAAGAGCCCCCACCTCAACTA
794400        34  GGTNGGANTCCGGGGNTGGGTAAAGACTTTAAGAGCCCCCACCTCAACTA
2850178       32  GGTCGCATCCCGGGGCTGGGTAAAGACTTTAAGAGCCCCCACCTCAACTA
1795703       20  GGTCGCATCCCGGGGCTGGGTAAAGACTTTAAGAGCCCCCACCTCAACTA
1809290        1          TCCCGGGGCTGGGTAAAGACTTTAAGAGCCCCCACCTCAACTA
1696222        1                    CGGCTCGANGGTCTCGAGTNACATNCNTNANT
1216292        1                                      CCCACCTCAACTA
                  ..+...+..++..++.++.+....++...+.........++.........
consen01    1097  GGTCGCATCCCGGGGCTGGGTAAAGACTTTAAGAGCCCCCACCTCAACTA

R24141       223  TTACTGGCTGCCCATCATGACCTCCATCCTGGGGGCCTATGTCATCGCC
2699329      221  TTACTGGCTGCCCATCATGACCTCCATCCTGGGGG-CCTATGTCATCGCC
3184889      205  TTACTGGCTGCCCATCATGACCTCCATCCTGGGGG-CCTATGTCATCGCC
816733       202  TTACTGGCTGACCATNANGACCTACAAACTGGNGN-ACTATGTCATGACA
1738719      165  TTACTGGCTGCCCATCATGACCTCCATCCTGGGGG-CCTATGTCATCGCC
1798567      149  TTACTGGCTGCCCATCATGACCTCCATCCTGGGGG-CCTATGTCATCGCC
2700889      142  TTACTGGCTGCCCATCATGACCTCCATCCTGGGGG-CCTATGTCATCGCC
```

FIGURE 3F

```
616205       140 TTACTGGNTGCCCNTCATGACCTNCATCCTGGGGG-CCTATGTCATCGGC
615116       138 TTACTGGNTGCCCATNATGANCTNCATNCTGGGGG-CCTATGTNATNGGC
1843426      121 TTACTGGCTGCCCATCATGACCTCCATCCTGGGGG-CCTATGTCATCGCC
794400        84 TTACTGGCTGCCCATCATGANCTCCATCCTGGGGG-CCTATGTCATCGNC
2850178       82 TTACTGGCTGCCCATCATGACCTCCATCCTGGGGG-CCTATGTCATCGCC
1795703       70 TTACTGGCTGCCCATCATGACCTCCATCCTGGGGG-CCTATGTCATCGCC
1809290       44 TTACTGGCTGCCCATCATGACCTCCATCCTGGGGG-CCTATGTCATCGCC
1696222       33 TCACTGGCTGCCCATCATGACCTCCATCCTGGGGG-CCTATGTCATCGCC
1216292       14 TTACTGGCTGCCCATCATGACCTCCATCCTGGGGG-CCTATGTCATCGCC
H25030         1   ACTNGCTGCCCATCATGACCTCCATCCTGGGGG-CCTATGTCATCGCC
1842150        1                     CCTCCATCCTGGGGG-CCTATGTCATCGCC
816972         1                          ATCCTGGGGG-CCTATGTCATCGCC
                 +.+++.+.++.++.+.+.++.++.++..++++.+. .++++++.++....
consen01    1147 TTACTGGCTGCCCATCATGACCTCCATCCTGGGGG CCTATGTCATCGCC

R24141       273 AGCGGTTNCTTCAGCGTTTTCGGGCATGTGTGTGGACACGCTCTTTCCT
2699329      270
3184889      254 AG-GG-TTCTTCAGCGTTT-CGG-CATGTGTGTGG-AA
816733       251 NCGGG
1738719      214 AG
1798567      198 AGCGGCTTCTTCAGCGTTTTCGG-CATGTGTGTGG-ACACGCTCTT-CCT
2700889      191 AGCGGCTTCTTCAGCGTTTTCGG-CATGTGTGTGG-ACACGCTCTT-CCT
616205       189 AGCGGGTTNTTCAGCGTTTNGG-GAATTGTGTGG-ACAAGGTNTT-CCT
615116       187 ANCGGNTTTTTNANCGTTTNGG-NATGTGTGTGG-ACAAGGT-TT-C-T
1843426      170 AGCGGCTTCTTCAGCGTTTTCGG-CATGTGTGTGG-ACACGCTCTT-CCT
794400       133 AGCGGCTTCTTCAGCGTTTTCGG-CATGTGTGTGG-ACAAGNTCTT-CCT
2850178      131 AGCGGCTTCTTCAGCGTTTTCGG-CATGTGTGTGG-ACACGCTCTT-CCT
1795703      119 AGCGGCTTCTTCAGCGTTTTCGG-CATGTGTGTGG-ACACGCTCTT-CCT
1809290       93 AGCGGCTTCTTCAGCGTTTTCGG-CATGTGTGTGG-ACACGCTCTT-CCT
1696222       82 AGCGGTTNCTTCAGCGTTTTCGG-CAT
1216292       63 AGCGGCTTCTTCAGCGTTTTCGG-CATGTGTGTGG-ACACGCTCTT-CCT
H25030        48 AGCGNTNTCTTCAGCGTTTTCGG-CATGTGTGTGG-ACACGCTCTT-CCT
1842150       30 AGCGG-TTCTTCAGCGTTTTCGG-CATGTGTGTGG-ACACGCTCTT-CCT
816972        25 AGCGGCTTCTTCAGCGTTTTCGG-CATGTGTGTGG-ACACGNTCTT-CCT
1736307        1                   CGGCTC-GAGGTGTGTGG-ACACGCTCTT-CCT
AA468365       1                      GG-CATGTGTGTGG-ACACGCTCTT-CCT
2274248.RC     1                                     AGACCTGTAA-GGC
                 ...+.....++.+.+++...... .+..+++++++ +.+....... ...
consen01    1196 AGCGGCTTCTTCAGCGTTTTCGG CATGTGTGTGG ACACGCTCTT CCT

R24141       323 CTGCTTCCTGGGGAAGGACCTGGGGAGCGTTAACAACGGGTTCCCTGGGG
1798567      245 CTGCTTCCTGG--AAG-ACCTGG--AGCGG-A
2700889      238 CTGCTTCCTGG--AAG-ACCTGG--A
616205       236 CTGCTTCCGGG--A--GACCTGG--AGCGG-AACAANGGNT-T
615116       232 CTGCTTC-TGG--GAG-ACCTGG--GGCGG-AA
1843426      217 CTGCTTCCTGG--AAG-ACCTGG--AGCGG-AACAACGGCT-CCCTGG--
794400       180 CTGCTTNCTGG--AAG-ANCTGG--AGCGG-AACAACGGCT-TCCTGG--
2850178      178 CTGCTTCCTGG--AAG-ACCTGG--AGCGG-AACAACGGCT-CCCTGG--
1795703      166 CTGCTTCCTGG--AAG-ACCTGG--AGCGG-AACAACGGCT-CCCTGG--
1809290      140 CTGCTTCCTGG--AAG-ACCTGG--AGCGG-AACAACGGCT-CCCTGG--
1216292      110 CTGCTTCCTGG--AAG-ACCTGG--AGCGG-AACAACGGCT-CCCTGG--
H25030        95 CTGCTTCCTGG--AAG-ACCTNG--AGCGG-AACAACGGCT-CCCTGG--
1842150       76 CTGCTTCCTGG--AAG-ACCTGG--AGCGG-AACAACGGCT-CCCTGG--
816972        72 CTGCTTCCTGG--AAG-ACCTGG--AGCGG-AACAANGGCT-CCCTGG--
1736307       31 CTGCTTCCTGG--AAG-ACCTGG--AGCGG-AACAACGGCT-CCCTGG--
AA468365      27 CTGCTTCCTGG--AAG-ACCTGG--AGCGG-AACAACGGCT-CCCTGG--
2274248.RC    14 GAAGTGAGGTT--GGATGGCTGG--AGCGG-AACAACGGCT-CCCTGG--
```

FIGURE 3G

```
                     ....+......  ......++.+  .+++. +++++.++.+ .+++++
consen01       1243 CTGCTTCCTGG  AAGGACCTGG  AGCGG AACAACGGCT CCCTGG

R24141          373 ACCGGCCCTACTNACATGTTCCAAGAGCCTTTTTAAAGATTCTGGGGCAA
1843426         258 ACCGG
794400          221 ACCGGCCCTACT-ACANGT-TCAAGA
2850178         219 ACCGGCCCTACT-ACATGT-CCAAGAGCCTTCT-AAAGATTCTGGG-CAA
1795703         207 ACCGGCCCTACT
1809290         181 ACCGGCCCTACT-ACATGT-CCAAGAGCCTTCT-AAAGATTCTGGG-CAA
1216292         151 ACCGGCCCTACT-ACATGT-CCAAGAGCCTTCT-AAAGATTCTGGG-CAA
H25030          136 ACCNGCCCTACT-ACATGT-CCAAGAGCCTTCT-AAAGATTCTGGG-CAA
1842150         117 ACCGGCCCTACT-ACATGT-CCAAGAGCCTTCT-AAAGATTCTGGG-CAA
816972          113 ACCGGCCCTACT-ACATGT-CCAAGAGCCTTCT-AAAGATTCTGGG-CAA
1736307          72 ACCGGCCCTACT-ACATGT-CCAAGAGCCTTCT-AAAGATTCTGGG-CAA
AA468365         68 ACCGGCCCTACT-ACATGT-CCAAGAGCCTTCT-AAAGATTCTGGG-CAA
2274248.RC       56 ACCGGCCCTACT-ACATGT-CCAAGAGCCTTCT-AAAGATTCTGGG-CAA
1431203           1                   GT-CCAAGAGCCTTCT-AAAGATTCTGGG-CAA
1862617           1                           GAGCCTTCT-AAAGATTCTGGG-CAA
842264            1                                    T-AAAGATTCTGGG-CAA
957891            1                                        AGATTCTGGG-CAA
1628246           1                                              GACTA-GTT
                    +++.++++++++ +++.++ .++++++++++.+ +++++++..... ...
consen01       1285 ACCGGCCCTACT ACATGT CCAAGAGCCTTCT AAAGATTCTGGG CAA

R24141          423 GAAGAACGAGGGCGCCCCCGGACAACAAGAAGAGG
2850178         265 GAAGAACGAGG-CGCCCCCGGACAACAAGAAGAGGAAGAAGTGACAG-TC
1809290         227 GAAGAAC
1216292         197 GAAGAACGAGG-CGCCCCCGGACAACAAGAAGAGGAAGAAGTGACAG-TC
H25030          182 GAAGAACGAGG-CGCCCCCGGACAACAAGAAGAGGAAGAAGTGACAGCTC
1842150         163 GAAGAACGAGG-CGCCCCCGGACAACAAGA
816972          159 GAAGAACGAGG-NGNCCCCGGACAACAAGAAGAGGAAGAAGTGACAGCTC
1736307         118 GAAGAACGAGG-CGCCCCCGGACAACAAGAAGAGGAAGAAGTGACAGCTC
AA468365        114 GAAGAACGAGG-CGCCCCCGGACAACAAGAAGAGGAAGAAGTGACAGCTC
2274248.RC      102 GAAGAACGAGG-CGCCCCCGGACAACAAGAAGAGGAAGAAGTGACAGCTC
1431203          31 GAAGAACGAGG-CGCCCCCGGACAACAAGAAGAGGAAGAAGTGACAGCTC
1862617          25 GAAGAACGAGG-CGCCCCCGGACAACAAGAAGAGGAAGAAGTGACAGCTC
842264           17 GAAGAACGAGG-CGCCCCCGGACAACAAGAAGAGGAAGAAGTGACAGCTC
957891           14 GAAGAACGAGG-CGCCCCCGGACAACAAGAAGAGGAAGAAGTGACAGCTC
1628246           9 CTAGAACGAGG-CGCCCCCGGACAACAAGAAGAGGAAGAAGTGACAGCTC
715308            1     AACGAGG-CGCCCCCGGACAACAAGAAGAGGAAGAAGTGACAGCNC
834390            1                               AGAAGAGGAAGAAGTGACAGCTC
2352555           1                                      GGAAGAAGTGACAGCTC
1820527           1                                              GTGACAGCTC
                    ..+++++++++ .+.+++++++++++++++++++++++++++++++.+
consen01       1331 GAAGAACGAGG CGCCCCCGGACAACAAGAAGAGGAAGAAGTGACAGCTC

2850178         313 CGGCCCTGATTCA
1216292         245 CGGCCC
H25030          231 CGGCCCTGATCCAGGACTGCACCCCACCCNCACCGTCCAGCCATCCAACC
816972          208 CGGCCCTGATCCAGGACTGCACCCCACCCCCACCGTCCAGCCATCCAACC
1736307         167 CGGCCCTGATCCAGGACTGCACCCCACCCCCACCGTCCAGCCATCCAACC
AA468365        163 CGGCCCTGATCCAGGACTGCACCCCACCCCCACCGTCCAGCCATCCAACC
2274248.RC      151 CGGCCCTGATCCAGGACTGCACCCCACC
1431203          80 CGGCCCTGATCCAGGACTGCACCCCACCCCCACCGTCCAGCCATCCAACC
1862617          74 CGGCCCTGATCCAGGACTGCACCCCACCCCCACCGTCCAGCCATCCAACC
842264           66 CGGCCCTGATCCAGGACTGCACCCCACCCCCACCGTCCAGCCATCCAACC
957891           63 CGGCCCTGATCCAGGACTGCACCCCACCCCCACCGTCCAGCCATCCAACC
```

FIGURE 3H

```
1628246          58 CGGCCCTGATCCAGGACTGCACCCCACCCCCACCGTCCAGCCATCCAACC
715308           46 CGGCCCTGNNCCAGGACNGCNCCCCACNCCCACCGTCCAGCCATCNAACC
834390           24 CGNNCCTNATCCAGGACTNCACCCCACCCNCACCGTNCAGCCATCCAACC
2352555          18 CGGCCCTGATCCAGGACTGCACCCCACCCCCACCGTCCAGCCATCCAACC
1820527          11 CGGCCCTGATCCAGGACTGCACCCCACCCCCACCGTCCAGCCATCCAACC
237332            1             CANGGACGTGCACCCCACCCCCACCGTCCAGCCATCCAACC
238940            1                CAGGACTGCACCCCACCCCCACCGTCCAGCCATCCAACC
2890547           1                                           CAGCCATCCAACC
                    ++..+++......+.....+.++++++.+.++++++.++++++++.++++
consen01       1380 CGGCCCTGATCCAGGACTGCACCCCACCCCCACCGTCCAGCCATCCAACC

H25030          281 TNACTTNGCCTTACAGGTCTCCATTTTGTGGTAAAAAAAGGTTTTAGGCC
816972          258 TNACT
1736307         217 TCACTTC
AA468365        213 TCACTTCGCCTTACAGGTCTCCATTTTGTGGTAAAAAAAGGTTTTAGGCC
1431203         130 TCACTTCGCCTTACAGGTCTCCATTTTGTGGTAAAAAAAGGTTTTA
1862617         124 TCACTTCGCCTTACAGGTCTCCATTTTGTGGTAAAAAAAGGTTTTA
842264          116 TCACTTCGCCTTACAGGTCTCCATTTTGTGGTAAAAAAAGGTTTTA
957891          113 TCACTTCGCCTTACAGGTCTCCATTTTGTGGTAAAAAAAGGTTTTA
1628246         108 TCACTTCGCCTTACAGGTCTCCATTTTGTGGTAAAAAAAGGTTTTA
715308           96 TCACTTCGCCTTACAGGTCTCCANTTTGTGGTAAAAANAGGTTTTA
834390           74 TCACTTCGNCTTACAGGTCTCCATTTTGTGGTAAAAAAAGGTTTTTG
2352555          68 TCACTTCGCCTTACAGGTCTCCATTTTGTGGTAAAAAAAGGTTTTA
1820527          61 TCACTTCGCCTTACAGGTCTCCATTTTGTGGTAAAAAAAGGTTTTA
237332           42 TCACTTCGCCTTACAGGTCTCCATTTTNTGGTAAAAAAAGGTTTTAGGCC
238940           40 TCACTTCGCCTTACAGGTCTCCATTTTGTGGTAAAAAAAGGTTTTA
2890547          14 TCACTTCGCCTTACAGGTCTCCATTTTGTGGTAAAAAAAGGTTTTA
                    +.++++.+.+++++++++++++.+++.+++++++++.+++++++.++++
consen01       1430 TCACTTCGCCTTACAGGTCTCCATTTTGTGGTAAAAAAAGGTTTTAGGCC

H25030          331 AGGGGNCGTGGCTTCACGCCTGTTAATTCAACACTTTGAGAGGNTTGAGG
AA468365        263 AGGCGCCGTGGCT-CACGCCTGT-AATCCAACACTTTGAGAGGCT-GAGG
237332           92 AGGCGCCGTGGCT-CACGCCTGT-AATCCAAC
                    +++.+.+++++++ +++++++++ +++.++++++++++++++.+ ++++
consen01       1480 AGGCGCCGTGGCT CACGCCTGT AATCCAACACTTTGAGAGGCT GAGG

H25030          381 CNGGCGGATTCAANTTGAGTTCAGGAGTTTCGAGANCAGTTGGGCCAACA
AA468365        310 CGGGCGGAT-CACCT-GAGT-CAGGAGTT-CGAGACCAGCCTGGCCAACA
                    +.+++++++ ++..+ ++++ ++++++++ +++++.+++...++++++++
consen01       1527 CGGGCGGAT CACCT GAGT CAGGAGTT CGAGACCAGCCTGGCCAACA

H25030          431 TTGTTGAAACTT
AA468365        356 TGGT-GAAACCTCCGTCTCTATTAAAAATACAAAAATTAGCCGAGAGTGG
                    +.++ +++++.++++++++++++++++++++++++++++++++++++++
consen01       1573 TGGT GAAACCTCCGTCTCTATTAAAAATACAAAAATTAGCCGAGAGTGG

AA468365        405 TGGCATGCACTGTCATCCAGCTACTCGGGAGGCTGAG
                    +++++++++++++++++++++++++++++++++++++
consen01       1622 TGGCATGCACTGTCATCCAGCTACTCGGGAGGCTGAG
```

FIGURE 4A

```
  1 GGGCCCCTGG TGCTGGTGTG ATCCTGGGAG TGCTGGGCGT GCTGGCATAT GGCATCTACT ACTGCTGGGA GGAGTACCGA GTGCTGCGGG ACAAGGGCGC
    CCCGGGGACC ACGACCACAC TAGGACCCTC ACGACCCGCA CGACCGTATA CCGTAGATGA TGACGACCCT CCTCATGGCT CACGACGCCC TGTTCCCGCG

101 CTCCATCTCC CAGCTGGGTT TCACCACCAA CCTCAGTGCC TACCAGAGCG TGCAGGAGAC CTGGCTGGCC GCCCTGATCG TGTTGGCGGT GCTTGAAGCC
    GAGGTAGAGG GTCGACCCAA AGTGGTGGTT GGAGTCACGG ATGGTCTCGC ACGTCCTCTG GACCGACCGG CGGGACTAGC ACAACCGCCA CGAACTTCGG

201 ATCCTGCTGC TGGTGCTCAT CTTCCTGCGG CAGCGGATTC GTATTGCCAT CGCCCTCCTG AAGGAGGCCA GCAAGGCTGT GGGACAGATG ATGTCTACCA
    TAGGACGACG ACCACGAGTA GAAGGACGCC GTCGCCTAAG CATAACGGTA GCGGGAGGAC TTCCTCCGGT CGTTCCGACA CCCTGTCTAC TACAGATGGT

301 TGTTCTACCC ACTGGTCACC TTTGTCCTCC TCCTCATCTG CATTGCCTAC TGGGCCATGA CTGCTCTGTA CCTGGCTACA TCGGGGCAAC CCCAGTATGT
    ACAAGATGGG TGACCAGTGG AAACAGGAGG AGGAGTAGAC GTAACGGATG ACCCGGTACT GACGAGACAT GGACCGATGT AGCCCCGTTG GGGTCATACA

401 GCTCTGGGCA TCCAACATCA GCTCCCCCGG CTGTGAGAAA GTGCCAATAA ATACATCATG CAACCCCACG GCCCACCTTG TGAACTCCTC GTGCCCAGGG
    CGAGACCCGT AGGTTGTAGT CGAGGGGGCC GACACTCTTT CACGGTTATT TATGTAGTAC GTTGGGGTGC CGGGTGGAAC ACTTGAGGAG CACGGGTCCC
                                                                                                     ^36459.f1

501 GCTGATGTGC GTCTTCCAGG GCTACTCATC CAAAGGCCTA ATCCAACGTT CTGTCTTCAA TCTGCAAATC TATGGGGTCC TGGGGCTCTT CTGGACCCTT
    CGACTACACG CAGAAGGTCC CGATGAGTAG GTTTCCGGAT TAGGTTGCAA GACAGAAGTT AGACGTTTAG ATACCCCAGG ACCCCGAGAA GACCTGGGAA
  1    M C   V F Q G   Y S S   K G L   I Q R S   V F N   L Q I   Y G V L   G L F   W T L
      ^orf                        ^36459.p1 .

601 AACTGGGTAC TGGCCCTGGG CCAATGCGTC CTCGCTGGAG CCTTTGCCTC CTTCTACTGG GCCTTCCACA AGCCCCAGGA CATCCCTACC TTCCCCTTAA
    TTGACCCATG ACCGGGACCC GGTTACGCAG GAGCGACCTC GGAAACGGAG GAAGATGACC CGGAAGGTGT TCGGGGTCCT GTAGGGATGG AAGGGGAATT
 33 N W V L   A L G   Q C V   L A G A   F A S   F Y W   A F H K   P Q D   I P T   F P L I

701 TCTCTGCCTT CATCCGCACA CTCCGTTACC ACACTGGGTC ATTGGCATTT GGAGCCCTCA TCCTGACCCT TGTGCAGATA GCCCGGGTCA TCTTGGAGTA
    AGAGACGGAA GTAGGCGTGT GAGGCAATGG TGTGACCCAG TAACCGTAAA CCTCGGGAGT AGGACTGGGA ACACGTCTAT CGGGCCCAGT AGAACCTCAT
 67   S A F   I R T   L R Y H   T G S   L A F   G A L I   L T L   V Q I   A R V I   L E Y

801 TATTGACCAC AAGCTCAGAG GAGTGCAGAA CCCTGTAGCC CGCTGCATCA TGTGCTGTTT CAAGTGCTGC CTCTGGTGTC TGGAAAAATT TATCAAGTTC
    ATAACTGGTG TTCGAGTCTC CTCACGTCTT GGGACATCGG GCGACGTAGT ACACGACAAA GTTCACGACG GAGACCACAG ACCTTTTTAA ATAGTTCAAG
100   I D H   K L R G   V Q N   P V A   R C I M   C C F   K C C   L W C L   E K F   I K F

901 CTAAACCGCA ATGCATACAT CATGATCGCC ATCTACGGGA AGAATTTCTG TGTCTCAGCC AAAAATGCGT TCATGCTACT CATGCGAAAC ATTGTCAGGG
    GATTTGGCGT TACGTATGTA GTACTAGCGG TAGATGCCCT TCTTAAAGAC ACAGAGTCGG TTTTTACGCA AGTACGATGA GTACGCTTTG TAACAGTCCC
133 L N R N   A Y I   M I A   I Y G K   N F C   V S A   K N A F   M L L   M R N   I V R V
```

FIGURE 4B

```
1001 TGGTCGTCCT GGACAAAGTC ACAGACCTGC TGCTGTTCTT TGGGAAGCTG CTGGTGGTCG GAGGCGTGGG GGTCCTGTCC TTCTTTTTTT TTCTCCGGTC
     ACCAGCAGGA CCTGTTTCAG TGTCTGGACG ACGACAAGAA ACCCTTCGAC GACCACCAGC CTCCGCACCC CCAGGACAGG AAGAAAAAAA AAGAGGCCAG
 167  V  V  L   D  K  V   T  D  L  L   L  F  F   G  K  L   L  V  V  G   G  V  G   V  L  S   F  F  F  F   L  R  S

1101 GCATCCCGGG GCTGGGTAAA GACTTTAAGA GCCCCCACCT CAACTATTAC TGGCTGCCCA TCATGACCTC CATCCTGGGG GCCTATGTCA TCGCCAGCGG
     CGTAGGGCCC CGACCCATTT CTGAAATTCT CGGGGGTGGA GTTGATAATG ACCGACGGGT AGTACTGGAG GTAGGACCCC CGGATACAGT AGCGGTCGCC
 200  H  P  G   A  G  O
                          ^36459.r1

1201 CTTCTTCAGC GTTTTCGGCA TGTGTGTGGA CACGCTCTTC CTCTGCTTCC TGGAAGGACC TGGAGCGGAA CAACGGCTCC CTGGACCGGC CCTACTACAT
     GAAGAAGTCG CAAAAGCCGT ACACACACCT GTGCGAGAAG GAGACGAAGG ACCTTCCTGG ACCTCGCCTT GTTGCCGAGG GACCTGGCCG GGATGATGTA

1301 GTCCAAGAGC CTTCTAAAGA TTCTGGGCAA GAAGAACGAG GCGCCCCCGG ACAACAAGAA GAGGAAGAAG TGACAGCTCC GGCCCTGATC CAGGACTGCA
     CAGGTTCTCG GAAGATTTCT AAGACCCGTT CTTCTTGCTC CGCGGGGGCC TGTTGTTCTT CTCCTTCTTC ACTGTCGAGG CCGGGACTAG GTCCTGACGT

1401 CCCCACCCCC ACCGTCCAGC CATCCAACCT CACTTCGCCT TACAGGTCTC CATTTTGTGG TAAAAAAAGG TTTTAGGCCA GGCGCCGTGG CTCACGCCTG
     GGGGTGGGGG TGGCAGGTCG GTAGGTTGGA GTGAAGCGGA ATGTCCAGAG GTAAAACACC ATTTTTTTCC AAAATCCGGT CCGCGGCACC GAGTGCGGAC

1501 TAATCCAACA CTTTGAGAGG CTGAGGCGGG CGGATCACCT GAGTCAGGAG TTCGAGACCA GCCTGGCCAA CATGGTGAAA CCTCCGTCTC TATTAAAAAT
     ATTAGGTTGT GAAACTCTCC GACTCCGCCC GCCTAGTGGA CTCAGTCCTC AAGCTCTGGT CGGACCGGTT GTACCACTTT GGAGGCAGAG ATAATTTTTA

1601 ACAAAAATTA GCCGAGAGTG GTGGCATGCA CTGTCATCCA GCTACTCGGG AGGCTGAG
     TGTTTTTAAT CGGCTCTCAC CACCGTACGT GACAGTAGGT CGATGAGCCC TCCGACTC
```